The Thesis committee for Shanny Lin
Certifies that this is the approved version of the following thesis:

A Spatio-Temporal Load Recovery Framework for Enhancing the Observability of Distributed Energy Resources

APPROVED BY

SUPERVISING COMMITTEE:

Hao Zhu, Supervisor

Surya Santoso

A Spatio-Temporal Load Recovery Framework for Enhancing the Observability of Distributed Energy Resources

by

Shanny Lin

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN ENGINEERING

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2020

Dedicated to my loving parents

for their endless love and support.

Acknowledgments

I would like to express my deep and sincere gratitude to my supervisor, Dr. Hao Zhu, for her guidance and support as well as for welcoming me into her research group. My achievements as a graduate student and this research project would not be possible without her patience, guidance, and encouragement. I would like to thank Dr. Surya Santoso for serving as a reader for this thesis and for his support. I would also like to thank The University of Texas at Austin Cockrell School of Engineering and the donors of the Virginia & Ernest Cockrell, Jr. Fellowship in Engineering. This fellowship support has allowed me to pursue exciting research experiences with few outside worries, and I am very lucky to have this opportunity. Finally, I would like to thank my family and friends for their unconditional love and support in my graduate studies.

Research reported in this thesis was supported by the National Science Foundation through Awards ECCS-1802319 and ECCS-1807097.

A Spatio-Temporal Load Recovery Framework for Enhancing the Observability of Distributed Energy Resources

Shanny Lin, M.S.E.
The University of Texas at Austin, 2020

Supervisor: Hao Zhu

This thesis presents algorithms that enhances the spatio-temporal observability of residential loads. This increased visibility in the load profiles is crucial for achieving secure and efficient operations in distribution systems, especially with the increasing penetration of distributed energy resources. The recovery formulation utilizes a joint inference framework by leveraging the spatial and temporal strengths of multiple data sources. More specifically, smart meter data is available for almost all residential households with low temporal resolution while distribution synchrophasor data is available at limited locations with very high temporal resolution. By combining the respective strengths of the two types of data, the load recovery problem is cast as a matrix recovery problem. Regularization terms are introduced to promote the underlying low-rank plus sparse structure characteristics of the load matrix to improve the recovery performance. As a result, the matrix recovery problem can be formulated as a convex optimization problem which can be solved using standard convex solvers. Numerical studies using real residential load

data demonstrate the effectiveness of the proposed algorithms capability in identifying large appliance activities and recovering the irradiance pattern of rooftop solar output profiles. Furthermore, our numerical studies have suggested that the presence of periodic loading can degrade the recovery performance. To address this issue, we have explored the introduction of an additional sinusoidal wave component. Last, online implementations of the recovery algorithms are discussed to accelerate the computational speed and process the data streams in real-time, while a rectangular waveform model is considered to better represent the presence of periodic loads. The proposed methods discussed in this thesis serve to enhance the observability of residential distributed energy resources.

Table of Contents

List of Figures

Chapter 1

Introduction

Power distribution systems are known to suffer from a lack of real-time observability, especially the residential household load profiles. This is due to the limited number of installed sensing devices, and thus sensor measurements. The available measurement data is typically obtained from a few line monitors and control devices. Traditional distribution state estimation (DSE) methods [2–4] utilize this limited amount of measurement data to improve system monitoring and control. However, these methods suffer from low estimation accuracy and robustness. Furthermore, they fail to improve the visibility of residential loads.

To address this challenge, advanced sensing, communication, and control technologies have been increasingly developed and deployed. One such sensing device is the distribution synchrophasor measurement unit (D-PMU) [5] which is

---

This chapter is based on the following conference paper:

1. Shanny Lin and Hao Zhu. "Enhancing the Spatio-Temporal Observability of Residential Loads." *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS 2020)*, Jan. 2020.

I am first author of this paper, and I am the main contributor to the load recovery formulation and the numerical studies.

installed at a few grid locations. D-PMUs provide synchronized high-quality phasor and power measurements with a fast time resolution, typically at sub-second sampling rates. In addition, smart meters are increasingly installed in every residential household to collect electricity consumption data at intervals of 15 minutes to one hour [6]. Notice that there exists a trade-off between the strengths of these two types of data. The smart meter data has high spatial diversity, however, it suffers from low temporal resolution as it loses transient information due to time averaging. Similarly, although D-PMUs have high sampling rates, they suffer from limited deployment due to high installation costs. As a result, neither of the sensing devices alone can provide the complete dynamics of the individual residential load profiles.

In addition to the lack of observability, the deeper penetration of distributed energy resources (DERs) poses a another challenge to the distribution system. The increasing penetration of DERs transforms the traditional single-source network to a multi-source one. As a result, enhanced spatio-temporal observability of residential loads is crucial in maintaining the distribution network security and stability. For example, the security of electric vehicle charging commands can be validated by estimating the start/end time of the charging period from the residential load profile with a high temporal resolution [7]. Similarly, the inverter control settings of residential photovoltaic (PV) systems can be verified using the dynamic load profile [8]. In addition, the increased solar generation in the network can significantly affect the fault current magnitude [9]. Therefore, the information contained in the PV output profiles is necessary to achieve effective distribution protection designs. Hence, the enhanced visibility of the residential load profiles are crucial

for achieving for achieving secure and reliable distribution system operations.

This thesis presents a joint inference framework to recover the residential real and reactive power load profiles by leveraging the strengths of both the smart meter and D-PMU data. In order to recover the spatio-temporal load matrices, the smart meter and aggregated power demand data collected by the D-PMU is first modeled as a linear transformation of the load matrices. To improve the recovery performance, the underlying characteristics of the load matrices are exploited. The real power load matrix has two key characteristics. It is observed that there exists a spatial correlation among the neighboring residential load nodes that is mainly due to PV generation and base loading. This spatial correlation can be represented by a *low-rank* component. In addition, there is an underlying *sparse-change* component that captures the individual household appliance activities at each residential node. [10] exploits this sparse-change characteristic to address the identifiability issue that occurs when only a subset of loads are metered. The reactive power load matrix, however, only has one key characteristic. The IEEE Standard 1547-2018 [11] recommends that PVs operate at unity power factor, and thus the reactive power matrix does not contain a low-rank component. However, it does contain a sparse-change component with entries that are *synchronized* with the sparse-change entries in the real power matrix as household appliances have a lagging power factor. To this end, two load recovery settings are considered here. The first one is a simplified load recovery model that assumes all real power loads are perfected aggregated with no line losses. Second, we consider the network power flow and line losses by incorporating both real and reactive power loads under the network

recovery model. The simplified recovery problem turns out to be recovering a low-rank plus sparse matrix. This problem formulation is very similar to the one used in robust principal component analysis (RPCA) works; see e.g., [12–14]. Note that similar low-rank plus sparse recovery approaches are popularly used in other power system applications such as synchrophasor data recovery [15, 16], load data cleansing [17], and forced oscillation localization [18]. However, the residential load matrices in the proposed framework requires a difference transformation to achieve the low-rank plus sparse structure. Similar to the earlier approaches, meaningful regularization terms are introduced to promote the two characteristics, and thus the simplified recovery problem can be cast as a convex optimization one. Similarly, the network recovery problem is obtained by extending the simplified model problem to include the recovery of the reactive power matrix. This network formulation would allow for numerical studies to be conducted on a lossy multiphase system.

The rest of this thesis is organized as follows. Chapter 2 introduces the smart meter and D-PMU measurement models. Chapter 3 introduces a simplified load recovery model that focuses only on recovering the real power matrix. The nuclear and L1-norm regularization are applied to promote the low-rank and sparse-change characteristics. Chapter 4 extends the simplified model formulation to include the network model and the recovery of the reactive power matrix. The nuclear norm is again utilized to promote the low-rank characteristic of the real power outputs from solar generation. Instead of the L1-norm, the group-Lasso regularization [19, 20] is applied to promote the sparse changes that simultaneously occur in both the real and reactive power matrices. Numerical tests using actual real power data demon-

strates the effectiveness of the simplified algorithm, while both actual real power data and generated reactive power data are used to demonstrate the effectiveness of the network algorithm. Note that it has been observed that the presence of periodic loading degrades the recovery performance. Chapter 5 introduces a sinusoidal wave component to address this challenge. The numerical results for a few of the individual load profiles show that the additional sinusoidal component attempts to capture the periodic loading. However, there is no visible improvement in regards to the low-rank component recovery. Chapter 6 concludes this thesis and discusses two future directions, namely online implementations of the recovery algorithms to accelerate the computational speed and to process data streams in real-time, while a rectangular waveform model is considered to better represent the presence of periodic loads.

Chapter 2

System Modeling

This thesis considers a generic unbalanced multiphase distribution feeder system with residential households connected at the feeder ends as shown in Fig. 2.1. Let N denote the number of single-phase loads (residential houses) and T the number of measurements taken at all loads over time. The goal is to recover the real and reactive spatio-temporal load matrices $\mathbf{P} \in \mathbb{R}^{N \times T}$ and $\mathbf{Q} \in \mathbb{R}^{N \times T}$ respectively. The temporal resolution of the load matrices represents the fastest sampling rate of all measurements. In this thesis, the highest resolution sampling rate is assumed to be every minute. However, the sampling rate can be generalized to even faster time-scales such as the sub-second time resolution of distribution phasor measurement units (D-PMUs) [5].

In order to recover the load matrices $\mathbf{P}$ and $\mathbf{Q}$, two types of measurement

---

This chapter is based on the following conference paper:

1. Shanny Lin and Hao Zhu. "Enhancing the Spatio-Temporal Observability of Residential Loads." *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS 2020)*, Jan. 2020.

I am first author of this paper, and I am the main contributor to the load recovery formulation and the numerical studies.

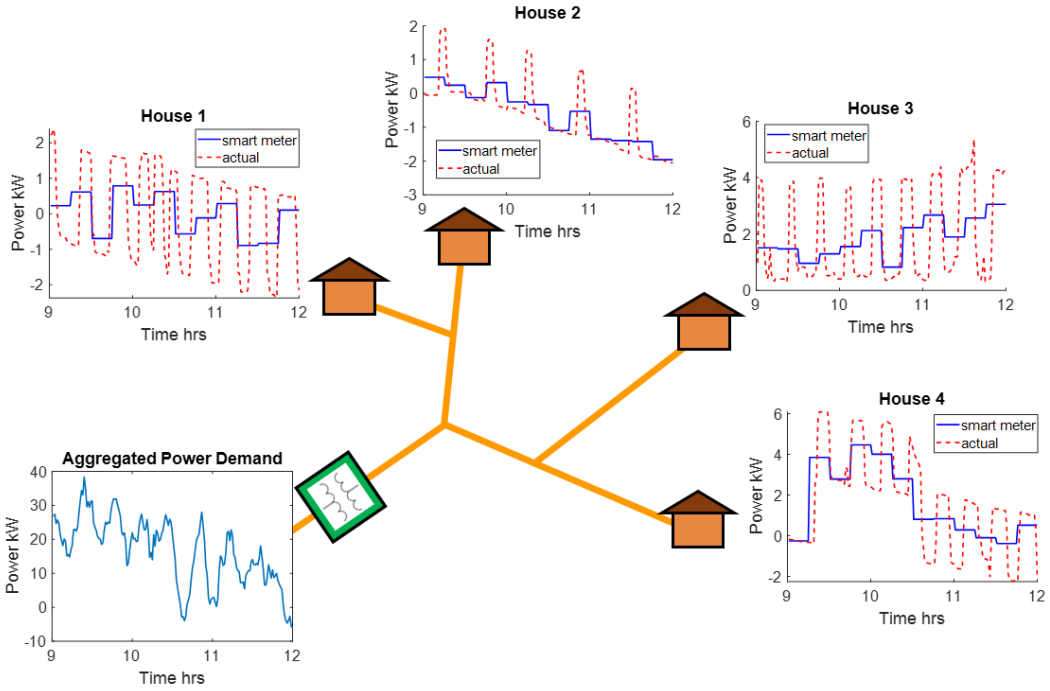


Figure 2.1: Overview of a simple distribution feeder system with four residential houses and various real power measurements.

data are utilized, namely i) the smart meter data at each load node and ii) the aggregated power demand at the feeder head. Note that in addition to the real and reactive demand profiles, the D-PMU installed at the feeder head also collects voltage and current phasor measurements with high quality and fast time resolution. The following sections will introduce the models defining the relationship between the smart meter and aggregated demand measurements to matrices $\mathbf{P}$ and $\mathbf{Q}$.

2.1. Smart Meter Data

Typically, smart meters are installed in every residential household to record the electricity consumption data over intervals of 15 minutes to an hour [6]. Note that in this thesis, smart meters are assumed to record both real and reactive power measurements. Additionally, it is assumed that these measurements are available every 15 minutes by averaging the corresponding 15 samples in $\mathbf{P}$ and $\mathbf{Q}$. Given this 1/15 down sampling rate, the smart meter real and reactive power data matrices $\mathbf{\Gamma}_p \in \mathbb{R}^{N\times T_s}$ and $\mathbf{\Gamma}_q \in \mathbb{R}^{N\times T_s}$, respectively, with $T_s = T/15$ is given by

$$\mathbf{\Gamma}_p = \mathbf{PA} + \mathbf{\Psi}_p \tag{2.1a}$$

$$\mathbf{\Gamma}_q = \mathbf{QA} + \mathbf{\Psi}_q \tag{2.1b}$$

where $\mathbf{A} = \mathbf{I}_{T_s} \otimes \mathbf{1}_{15\times 1} \in \mathbb{R}^{T\times T_s}$ is the time averaging operator, and $\mathbf{\Psi}_p$ and $\mathbf{\Psi}_q$ denote the real and reactive measurement noise matrices respectively.

2.2. Power Demand Data

This section defines the model of the aggregated demand seen at the feeder head. In order to formulate the demand model, the multiphase AC power flow equations will be briefly introduced; see [21] and [22] for a more comprehensive description. To better visualize the AC power flow variables on the distribution feeder system, let the slack bus be the node at the feeder head and N_b denote the number of PQ buses. Then, the set $\mathcal{N} := \{0, ..., N_b\}$ defines the collection of buses in the system where bus 0 indicates the slack bus the slack bus. Without loss of generality, all buses are three-phase connected. At each multiphase node, the loads

can be connected in either wye or delta configurations [22]. For simplicity, all loads considered in this thesis are assumed to be wye-connected. This assumption holds true since the residential homes were previously defined as single-phase loads. To this end, let $\mathbf{s}_j = [s_j^a, s_j^b, s_j^c]^\mathsf{T}$ denoted the complex power injections of at bus $j \in \mathcal{N}$. Similarly, let $\mathbf{v}_j = [v_j^a, v_j^b, v_j^c]^\mathsf{T}$ and $\mathbf{i}_j = [i_j^a, i_j^b, i_j^c]^\mathsf{T}$ denote the phase-to-ground voltages and the phase current injections at bus $j \in \mathcal{N}$ respectively. The complex vectors of length $3N_b$ collecting the quantities for all N_b PQ buses can be defined as $\mathbf{s} = [\mathbf{s}_1{}^\mathsf{T}, ..., \mathbf{s}_{N_b}{}^\mathsf{T}]^\mathsf{T}$,$\mathbf{v} = [\mathbf{v}_1{}^\mathsf{T}, ..., \mathbf{v}_{N_b}{}^\mathsf{T}]^\mathsf{T}$, and $\mathbf{i} = [\mathbf{i}_1{}^\mathsf{T}, ..., \mathbf{i}_{N_b}{}^\mathsf{T}]^\mathsf{T}$. Hence, the multiphase AC power flow equations are defined as

$$\mathbf{s} = \mathsf{diag}(\mathbf{v})\mathbf{i}^* \tag{2.2a}$$

$$\mathbf{i} = \mathbf{Y}_{L0}\mathbf{v}_0 + \mathbf{Y}_{LL}\mathbf{v} \tag{2.2b}$$

where the submatrices $\mathbf{Y}_{L0} \in \mathbb{C}^{3N_b \times 3}$ and $\mathbf{Y}_{LL} \in \mathbb{C}^{3N_b \times 3N_b}$ corresponding to the non-slack buses are extracted from the full three-phase admittance matrix

$$\mathbf{Y} := \begin{bmatrix} \mathbf{Y}_{00} & \mathbf{Y}_{0L} \\ \mathbf{Y}_{L0} & \mathbf{Y}_{LL} \end{bmatrix} \in \mathbb{C}^{3(N_b+1)\times 3(N_b+1)}; \tag{2.3}$$

see [21] for complete details. Similarly, the slack bus complex power injection (aggregated demand) is given by

$$\mathbf{s}_0 = \mathsf{diag}(\mathbf{v}_0)(\mathbf{Y}_{00}^*\mathbf{v}_0^* + \mathbf{Y}_{0L}^*\mathbf{v}^*) \in \mathbb{C}^3 \tag{2.4}$$

where $\mathbf{v}_0$ denotes the slack bus voltage. Assuming the slack bus has constant voltage $\mathbf{v}_0$, the aggregated demand $\mathbf{s}_0$ is fully determined by the feeder voltage profile in vector $\mathbf{v}$.

To efficiently solve for $\mathbf{v}$, [21] proposed a linearized AC power flow model by using a fixed-point approximation of (2.2) given by

$$\tilde{\mathbf{v}} = \mathbf{M}\mathbf{x} + \mathbf{w} \tag{2.5}$$

where the real and reactive demand is concatenated into $\mathbf{x} \coloneqq [\mathsf{Re}\{\mathbf{s}\}^\mathsf{T}, \mathsf{Im}\{\mathbf{s}\}^\mathsf{T}]^\mathsf{T}$ and $\mathbf{w}$ is the zero-load voltage. The matrix $\mathbf{M}$ is computed based on the network parameters and a given voltage profile $\hat{\mathbf{v}}$:

$$\mathbf{M} \coloneqq [\mathbf{Y}_{LL}^{-1}\mathsf{diag}(\hat{\mathbf{v}}^*)^{-1}, -j\mathbf{Y}_{LL}^{-1}\mathsf{diag}(\hat{\mathbf{v}}^*)^{-1}]. \tag{2.6}$$

This fixed-point linearization method has been considered in other papers [23–25] and can be generalized to include delta-connected loads as shown in [21]. Although not considered in this thesis, (2.2) and (2.5) can be generalized to include delta-connected loads. Plugging (2.5) into (2.4) gives rise to the linear aggregated power demand as defined by

$$\mathbf{s}_0 = \mathsf{diag}(\mathbf{v}_0)(\mathbf{Y}_{00}^*\mathbf{v}_0^* + \mathbf{Y}_{0L}^*\tilde{\mathbf{v}}^*) \tag{2.7a}$$

$$= \mathsf{diag}(\mathbf{v}_0)\mathbf{Y}_{0L}^*\mathbf{M}^*\mathbf{x} \tag{2.7b}$$

$$= \mathbf{H}\mathbf{x} \tag{2.7c}$$

where $\mathbf{H} = \mathsf{diag}(\mathbf{v}_0)\mathbf{Y}_{0L}^*\mathbf{M}^*$ denotes the transformation matrix on $\mathbf{x}$. Note that the other term $\mathsf{diag}(\mathbf{v}_0)(\mathbf{Y}_{00}^*\mathbf{v}_0^* + \mathbf{Y}_{0L}^*\mathbf{w}^*)$ is a constant and actually equals to zero as it defines power injection under zero-loading conditions, namely $\mathbf{x} = \mathbf{0}$. Furthermore, stacking (2.7) across T time slots leads to the aggregated demand matrix $\mathbf{Z} = \mathbf{H}\mathbf{X} \in \mathbb{C}^{3\times T}$, and similarly one can define matrix $\mathbf{X}$ to collect the corresponding real and reactive power injection vectors. Note that $\mathbf{Z} = \mathbf{H}\mathbf{X}$ is consistent

with (2.8). It can observed that the dimensions of both $\mathbf{H}$ and $\mathbf{X}$ can be reduced. By eliminating the entries for no-loading buses in $\mathbf{X}$ and corresponding columns in $\mathbf{H}$, one can reduce the dimension to match the number of loads N as defined earlier. With slight misuse of notation, the rest of the thesis will adopt this reduced version of $\mathbf{H}$ and $\mathbf{X}$. Hence, the matrix $\mathbf{X} = -[\mathbf{P}^\mathsf{T}, \mathbf{Q}^\mathsf{T}]^\mathsf{T}$ directly relates to the aforementioned load matrices, and the measurement matrix $\mathbf{Z}$ is given by

$$\mathbf{Z} = \mathbf{H}\mathbf{X} + \mathbf{\Theta} \tag{2.8}$$

where $\mathbf{\Theta}$ is the D-PMU measurement noise matrix. The negative sign is introduced here to be consistent with the load demands $\mathbf{P}$ and $\mathbf{Q}$. One can further rewrite (2.8) with the respective real and reactive power matrices given by

$$\mathbf{Z}_p = \mathbf{H}_p\mathbf{X} + \mathbf{\Theta}_p \tag{2.9a}$$

$$\mathbf{Z}_q = \mathbf{H}_q\mathbf{X} + \mathbf{\Theta}_q \tag{2.9b}$$

where $\mathbf{H}_p = \mathsf{Re}\{\mathbf{H}'\}$ and $\mathbf{H}_q = \mathsf{Im}\{\mathbf{H}\}$.

Clearly, using the measurements defined by (2.1) and (2.8), the problem of recovering the load matrices $\mathbf{P}$ and $\mathbf{Q}$ is under-determined. With a very large number of loads N, the total number of equations, given by (2.1) and (2.8), equals $(2NT_s + 6T)$ which is much smaller than $2NT$, the number of unknowns. Interestingly, both matrices $\mathbf{P}$ and $\mathbf{Q}$ have special underlying characteristics and structures that will be exploited to tackle the low measurement issue for improved recovery performance.

Chapter 3

Simplified Load Recovery Model

This chapter adopts a simplified aggregated demand model to illustrate the recovery of the real power demand $\mathbf{P}$. The distribution network presented previously in Chapter 2 is assumed to be a lossless single-phase network. However, note that this assumption can be generalized to include minimal or constant losses. By utilizing this simplified network, the effectiveness of exploiting the underlying characteristics of $\mathbf{P}$ in load recovery can be shown. In addition, oftentimes the real power measurements $\mathbf{P}$ is available, but not for the reactive power $\mathbf{Q}$. Hence, the algorithm introduced later in this chapter only takes in real power smart meter and D-PMU measurements as inputs.

To this end, only the real power smart meter and aggregated demand measurements, $\mathbf{\Gamma}_p$ and $\mathbf{Z}_p$, will be considered. However, the aggregated demand model

---

This chapter is based on the following conference paper:

1. Shanny Lin and Hao Zhu. "Enhancing the Spatio-Temporal Observability of Residential Loads." *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS 2020)*, Jan. 2020.

I am first author of this paper, and I am the main contributor to the load recovery formulation and the numerical studies.

(2.9a) is reduced to a vector

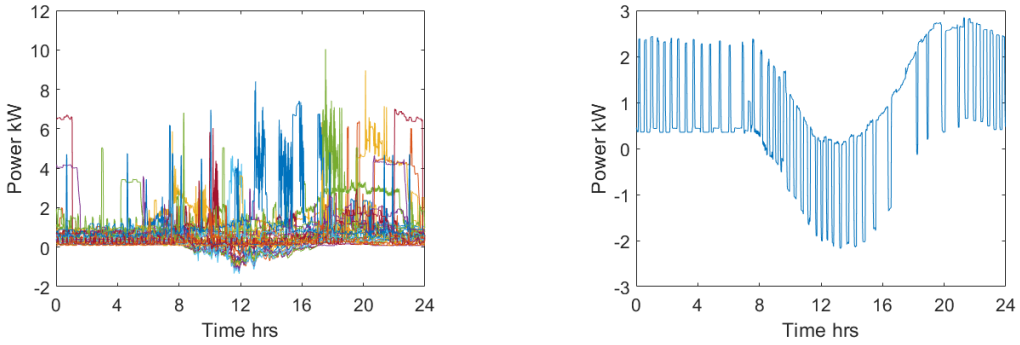
$$\mathbf{z}_p^\mathsf{T} = \mathbf{1}^\mathsf{T}\mathbf{P} + \boldsymbol{\theta}_p^\mathsf{T} \tag{3.1}$$

as the network is assumed to be single-phase and lossless. Additionally, note that this simplified system of equations is still under-determined as the total number of equations is $(NT_s + T)$ and the number of unknowns is NT.

3.1. Matrix $\mathbf{P}$ Characteristics

Transmission level load demand curves have been extensively studied throughout the years. As a result, it is well known that the transmission level load curves exhibit a high spatial correlation among neighboring locations and, therefore, these locations will share similar *temporal patterns*. The spatial correlation property is widely exploited in load forecasting and data cleansing works; see e.g. [17, 26–28]. The temporal pattern results from underlying characteristics such as weather conditions (e.g. temperature and solar irradiace) and economic conditions (i.e. electricity prices). Although this property holds for transmission level loads, it is observed to hold true for distribution level loads as well.

To better observe this property in the distribution level, actual residential load demand profiles are retrieved from Pecan Street's Dataport [1] with minute-level time resolution. In addition, the selected residential houses are located in the same neighborhood in Austin, Texas. Fig. 3.1a plots the load profiles of 30 houses, 15 of which have photovoltaics (PVs), for one winter day and Fig. 3.1b plots the load profile of one selected house with PV for one summer day. Compared to the



(a) (b)

Figure 3.1: Actual residential load profiles retrieved from Pecan Street's Dataport [1] for (a) 30 houses for one winter day and (b) a single house with PV for one summer day.

winter loading profiles, the summer load profile has high levels of heating, ventilation, and air conditioning (HVAC) activity. From Fig. 3.1a, it can be observed that there exists a daytime temporal pattern among the 15 houses with PVs, and thus they share a similar solar irradiance pattern. Similarly, the 15 houses without PVs share a minimal base loading pattern (loading without electrical appliance activity). Additionally, all 30 houses share a similar base loading pattern during time periods without solar irradiance. Therefore, the spatial correlation among the minute-level residential load profiles mainly results from the PV generation output and base loading. This spatial correlation in $\mathbf{P}$ is represented by an underlying low-rank component $\mathbf{L} \in \mathbb{R}^{N \times T}$ with rows that are either highly correlated (loads with PVs) or close to zero (loads without PVs).

Beyond the spatial correlation, distribution load curves are also observed to contain rectangular waveforms that are not synchronized across loads. These wave-

forms correspond to large appliance activities in individual households (e.g. EV charging and HVAC). For example, the visible rectangular pulses in Fig. 3.1a correspond to EV charging events and the periodic rectangular waveform in Fig. 3.1b correspond to HVAC loads. In general, these large appliance events are observed to occur infrequently throughout the day and are not synchronized across households. As a result, these events can be thought of as *sparse changes* and can be captured with an additional component $\mathbf{S} \in \mathbb{R}^{N \times T}$ in load matrix $\mathbf{P}$. Note that $\mathbf{S}$ itself is piece-constant, and thus the consecutive differences in the entries of $\mathbf{S}$ are sparse, as defined by $D_{n,t} = S_{n,t} - S_{n,t-1} \forall (n,t)$ with $S_{n,t} = \sum_{\tau=1}^{t} D_{n,\tau}$. Therefore, $\mathbf{S}$ becomes sparse under the linear transformation given by $\mathbf{S} = \mathbf{DU}$ where $\mathbf{U} \in \mathbb{R}^{T \times T}$ is an upper triangular matrix of all ones. As a result, the load matrix $\mathbf{P}$ can be decomposed into a low-rank matrix augmented with a sparse change matrix such that $\mathbf{P} = \mathbf{L} + \mathbf{S} = \mathbf{L} + \mathbf{DU}$.

3.2. Recovery Formulation

This decomposition of $\mathbf{P}$ is very similar to the model framework used in robust principal component analysis (RPCA), which decomposes a large data matrix into the sum of a low-rank and sparse component [12]. In RPCA, the low-rank component also captures the spatial correlation among rows and the sparse component captures outliers that are not spatially correlated. Hence, the decomposition of $\mathbf{P}$ is slightly different than the RPCA framework. However, by introducing a linear transformation matrix $\mathbf{W} = \mathbf{U}^{-1}$, which computes the consecutive differences

between columns, the transformed model of $\mathbf{P}$ is given as

$$\mathbf{R} \coloneqq \mathbf{PW} = (\mathbf{L} + \mathbf{DU})\mathbf{W} = \mathbf{K} + \mathbf{D} \tag{3.2}$$

where $\mathbf{K} = \mathbf{LW}$ is also a low-rank matrix. Clearly, the decomposition of the transformed matrix $\mathbf{R}$ now satisfies the RPCA's low rank plus sparse structure. Relating the new $\mathbf{K}$ and $\mathbf{D}$ components back to the measurement models in (2.1a) and (3.1), the updated models are given as

$$\mathbf{\Gamma}_p = (\mathbf{K} + \mathbf{D})\mathbf{UA} + \mathbf{\Psi}_p \tag{3.3a}$$

$$\mathbf{z}_p^\mathsf{T}\mathbf{W} = \mathbf{1}^\mathsf{T}(\mathbf{K} + \mathbf{D}) + \boldsymbol{\theta}_p^\mathsf{T}\mathbf{W}. \tag{3.3b}$$

The problem of recovering load matrix $\mathbf{P}$ now becomes one of recovering a low-rank plus sparse matrix. To promote this special structure in the recovery process, meaningful regularization terms are introduced. The matrix nuclear norm is widely adopted in low-rank matrix recovery problems such as matrix completion [12], subspace learning [14], and collaborative filtering [29]. The nuclear norm is defined as

$$\|\mathbf{K}\|_* \coloneqq \sum_{i=1}^{\min\{N,T\}} \sigma_i(\mathbf{K}) \tag{3.4}$$

where $\sigma_i(\cdot)$ denotes the *i*-th largest singular value. In addition, the nuclear norm is a convex norm since it is the dual function of the matrix spectral norm (L2-norm); see [30, pg. 637] for details.

To promote the sparse matrix $\mathbf{D}$, the popular L1-norm regularization is adopted, as used in the fields of compressed sensing and sparse signal recovery;

see e.g. [31–33]. The L1-norm is defined as the sum of the entry-wise absolute values, given by

$$\|\mathbf{D}\|_1 \coloneqq \sum_{n,t} |\mathbf{D}_{n,t}|. \tag{3.5}$$

The L1-norm is the tight convex relaxation of the L0 pseudo-norm, and thus it has been shown to be capable of efficiently finding sparse signal representations with performance guarantees. Additionally, note that the nuclear norm defined in (3.4) is essentially an L1-norm on matrix $\mathbf{K}$'s singular values. Therefore, minimizing the nuclear norm will reduce the number of non-zero singular values which effectively gives a low-rank matrix solution.

Utilizing the nuclear norm and L1-norm regularizations, the matrix $\mathbf{R}$ recovery problem can be defined as

$$\min_{\mathbf{K},\mathbf{D}} \quad \|\mathbf{K}\|_* + \lambda\|\mathbf{D}\|_1 \tag{3.6a}$$

$$\text{s. to} \quad -\boldsymbol{\xi} \leq \boldsymbol{\Gamma}_p - (\mathbf{K}+\mathbf{D})\mathbf{U}\mathbf{A} \leq \boldsymbol{\xi} \tag{3.6b}$$

$$-\boldsymbol{\phi} \leq \mathbf{z}_p^{\mathsf{T}}\mathbf{W} - \mathbf{1}^{\mathsf{T}}(\mathbf{K}+\mathbf{D}) \leq \boldsymbol{\phi} \tag{3.6c}$$

where $\lambda > 0$ is a fixed weight coefficient to balance matrices $\mathbf{K}$ and $\mathbf{D}$ and the parameters $\boldsymbol{\xi}, \boldsymbol{\phi} > 0$ are pre-determined entry-wise error bounds. These error bounds can be set according to the metering accuracy of each type of measurement. Furthermore, it can even include potential modeling inaccuracy due to e.g., bad data or feeder losses. If only measurement noise is considered, the error constraints in (3.6b)-(3.6c) correspond to uniformly distributed noise and the error is bound using the infinity norm. As such, both $\boldsymbol{\xi}$ and $\boldsymbol{\phi}$ are entry-wise error bounds. This assumption holds true in practical systems since the metering accuracy is specified by

a maximum error percentage. As a result, different entry-wise values are allowed in the error bounds, since the error will scale with the corresponding measurement entry. Additionally, different error criteria can be used to replace (3.6b)-(3.6c) such as the Frobenius norm for Gaussian distributed noise. Since the recovery problem in (3.6) has a convex objective and linear constraints, it can be solved using generic convex optimization solvers such as CVX [34] and YALMIP [35]. The computational complexity, however, will grow with the matrix dimension. This issue can be overcome by utilizing accelerated algorithms such as alternating minimization or adaptive updates using subspace learning approaches, but these approaches are out of scope for this thesis.

There is, however, one main issue with the formulation presented in (3.6). Due to the L1-norm regularization, the magnitudes on the nonzero entries of $\mathbf{D}$ may be penalized which results in a biased solution. This effect can be especially observed in the summer numerical studies in 3.3.2 where the presence of periodic HVAC activities could exacerbate this issue. The periodicity of the HVAC activity can exhibit a certain level of temporal pattern which can be partially captured by $\mathbf{L}$. Therefore, the entries in the recovered $\hat{\mathbf{D}}$ will potentially suffer from smaller magnitudes.

To overcome this issue, a post-processing scheme was developed based on the recovered support of $\hat{\mathbf{D}}$. Despite the existence of the magnitude bias, (3.6) can accurately determine the locations of the nonzero entries of $\mathbf{D}$. Therefore, it is possible to build a support set $\mathcal{M} = \{(n,t)\}|\ |\hat{D}_{n,t}| > 0\}$ using $\hat{\mathbf{D}}$. Note that in the numerical tests, a small positive threshold is used to reflect the numerical

Algorithm 1 Recovering matrices $\mathbf{K}$ and $\mathbf{D}$

1: **Input:** Smart meter data $\mathbf{\Gamma}_p$ and D-PMU data $\mathbf{z}_p$.
2: **Output:** Estimated $\hat{\mathbf{K}}$ and $\hat{\mathbf{D}}$
3: **Step 1:** Solve the problem (3.6) to obtain the biased estimation $\hat{\mathbf{K}}^b$ and $\hat{\mathbf{D}}^b$.
4: **Step 2:** Find the set of nonzero entries $\mathcal{M} := \{(n,t)|\ |\hat{D}_{n,t}| > 0\}$.
5: **Step 3:** Solve the problem (3.7) using $\mathcal{M}$ to obtain the updated estimates $\hat{\mathbf{K}}$ and $\hat{\mathbf{D}}$.

accuracy of the zero entries. Utilizing the support set $\mathcal{M}$ as an additional constraint, the L1-norm regularization can be neglected to recast (3.6) as

$$\min_{\mathbf{K},\mathbf{D}} \quad \|\mathbf{K}\|_* \tag{3.7a}$$

$$\text{s. to} \quad D_{n,t} = 0, \forall (n,t) \notin \mathcal{M} \tag{3.7b}$$

$$-\boldsymbol{\xi} \leq \mathbf{\Gamma}_p - (\mathbf{K}+\mathbf{D})\mathbf{U}\mathbf{A} \leq \boldsymbol{\xi} \tag{3.7c}$$

$$-\boldsymbol{\phi} \leq \mathbf{z}_p^{\mathsf{T}}\mathbf{W} - \mathbf{1}^{\mathsf{T}}(\mathbf{K}+\mathbf{D}) \leq \boldsymbol{\phi}. \tag{3.7d}$$

This new post-processing problem (3.7) allows for better estimated values for the nonzero entries of $\mathbf{D}$. The full recovery formulation for $\mathbf{K}$ and $\mathbf{D}$ is tabulated in Algorithm 1.

3.2.1 Recovery Performance

This subsection discusses the performance guarantees of the recovery formulation (3.6). As mentioned previously, the framework of this thesis is similar to that of RPCA's. As a result, the regularization terms on the low-rank and sparse components are the same which results in the same objective function. However, the RPCA framework assumes that the full data matrix itself is either fully or partially observable [12–14]. In this thesis, the unknown data matrix $\mathbf{P}$ is observed

with reduced dimensionality in both column and row space at the ratios of T/T_s and $1/N$ respectively. Therefore, the theoretical results proven in RPCA are not directly applicable to the framework of this thesis. However, RPCA results can provide intuition on the recovery conditions and the weight parameter λ settings in (3.6).

The RPCA framework assumes specific conditions on the low-rank and sparse components, $\mathbf{K}$ and $\mathbf{D}$ respectively, to achieve good recovery results [12, 36, 37]. It is enforced that the low-rank component $\mathbf{K}$ cannot be sparse. Similarly, the sparse component $\mathbf{D}$ cannot be low-rank. Thus, for $\mathbf{K}$, the column spanned by either its left or right singular vectors needs to have low or zero coherency with the identity matrix and, thereby, ensuring the singular vectors are not sparse. In relation to this thesis, it implies that temporal pattern of the load profiles itself cannot be sparse. In regards to $\mathbf{D}$, the locations of the sparse non-zero entries should be random with no inherent periodic or correlated patterns. As a result, $\mathbf{D}$ would be ideally full rank and the non-periodic condition ensures that no frequency induced low-rank component exits in $\mathbf{D}$. This non-periodic condition is crucial to promote the ideal separation of the low-rank and sparse components. As shown later in the simulation results on the summer dataset in Sec. 3.3.2, the high level of periodic HVAC activities degrades the quality of the recovery results.

The other insight provide by RPCA is the weight parameter λ setting that is used to balance the low-rank and sparse components. As discussed in [12, 14], the λ value is dependent on the matrix dimension. For example, if $T \gg N$, then the λ value is chosen as $\mathcal{O}(1/\sqrt{T})$. This setting balances the nuclear and L1-norm with

respect to T, at around $\sqrt{T}$ and T respectively.

3.3. Numerical Tests

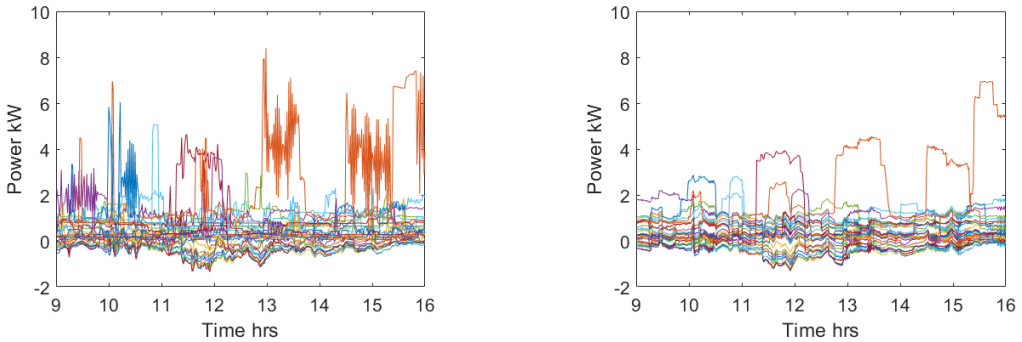
Numerical studies using Algorithm 1 to recover $\mathbf{P}$ was conducted on a winter and summer dataset. The ground-truth $\mathbf{P}$ for each dataset was obtained from Pecan Street's Dataport [1] for 30 residential homes in the same neighborhood (Mueller, Austin, TX); the winter dataset is shown in Fig. 3.1a. The time resolution of the datasets are at the minute-level. Additionally, only half of the houses have PVs and 6 of the houses have EV charging events. The smart meter and aggregated demand measurements, (2.1a) and (3.1) respectively, were synthetically generated by adding random noise. Note that the smart meters are assumed to record data at 15 minute intervals. From the American National Standard Institute (ANSI) C12.20 Standard [38], the smart meters are assumed to be rated with $\pm 0.2\%$ accuracy. Therefore, the entries of the smart meter measurement noise matrix $\mathbf{\Psi}_p$ are independently drawn from a uniform distribution based on this accuracy. Similarly using the D-PMU data-sheet [39] for the aggregated demand measurement $\mathbf{z}_p$, the $\pm 0.01\%$ total vector error rating on phasor measurements results in $\pm 0.02\%$ error in active power measurements. Hence, the entries of the aggregated measurement noise vector $\boldsymbol{\theta}_p$ are independently drawn from uniform distributions using $\pm 0.02\%$ accuracy.

In addition to the measurement noise, the values of the weight coefficent λ and the error bounds, $\boldsymbol{\xi}$ and ϕ, need to be determined. In all of the following numerical studies, the weighting parameter λ was fixed at a value of 0.05. This value was

chosen based upon $\mathcal{O}(1/\sqrt{T})$ as discussed in Sec. 3.2.1. The error bounds $\boldsymbol{\xi}$ and $\boldsymbol{\phi}$ were determined based upon the metering accuracy used to generate the synthetic data. Hence, ξ_{y_p} and ξ_{z_p} are set to $0.2\%|\mathbf{Y}|$ and $0.02\%|\mathbf{z}|^{\mathsf{T}}|\mathbf{W}|$ respectively.

3.3.1 Winter Data Results

The recovery formulation (3.6) is tested on two time periods in the winter data set, namely one day-time period and one night-time period. The day-time period from 9:00-16:00 was tested first to recover the solar irradiance pattern. Starting with the overall recovery of $\mathbf{P}$, Fig. 3.2b plots the estimated $\hat{\mathbf{P}}$ which well matches the trends in the ground-truth profiles shown in Fig. 3.2a. Although $\hat{\mathbf{P}}$ is unable to recover the fast transients, it is capable of capturing the major dynamic changes in the load profiles such as the large household appliance events and PV dynamics. To better observe the solar irradiance pattern recovery, Fig. 3.3 shows the comparison between the estimated $\hat{\mathbf{L}}$ and the ground-truth PV output profiles. Note that the



(a) (b)

Figure 3.2: Comparison between (a) the ground-truth residential load profiles and (b) the recovered $\hat{\mathbf{P}}$ from (3.6) for the winter day-time period.

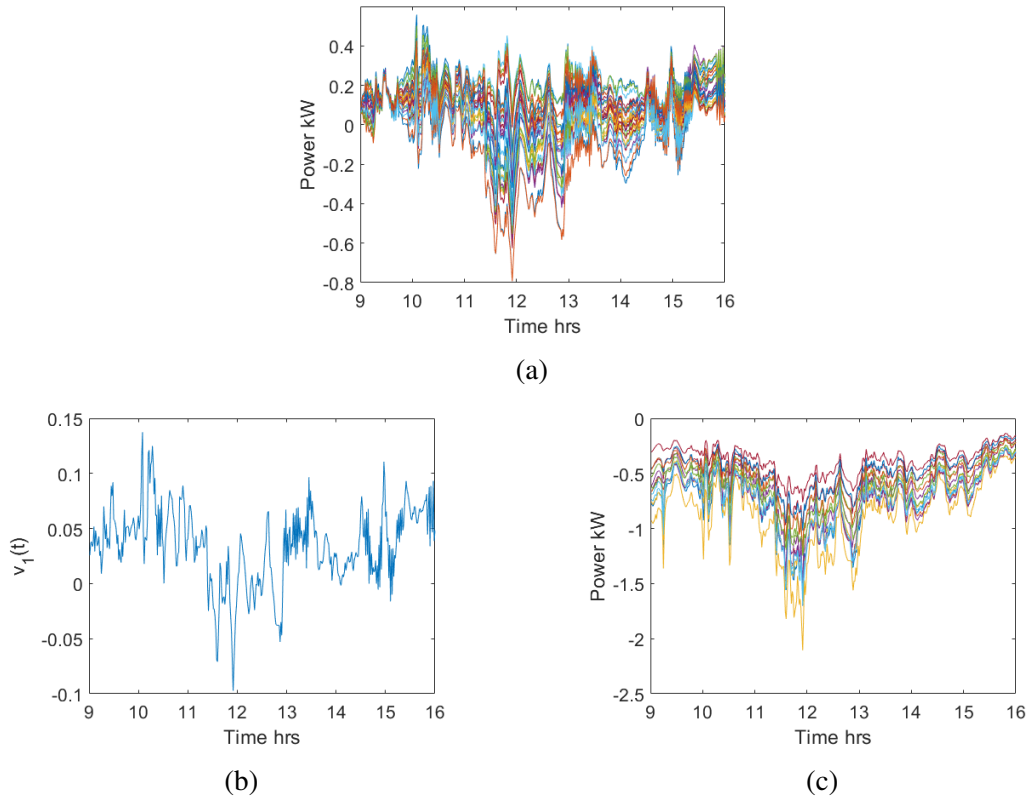


Figure 3.3: Comparison of the winter day-time solar irradiance pattern recovery of (a) the estimated $\hat{\mathbf{L}}$ and (b) its first right singular vector with (c) the ground-truth PV output profile.

estimated $\hat{\mathbf{L}}$ shown in Fig. 3.3a does not recover the actual PV outputs in Fig. 3.3c. However, the estimated $\hat{\mathbf{L}}$ clearly has very low rank as the majority of its temporal pattern is reflected in its first right singular vector shown in Fig. 3.3b. In addition, the solar irradiance pattern captured by the first right singular vector well matches the pattern present in the ground-truth PV outputs. Compared to the summer load profiles, the winter dataset does not contain HVAC activities. Hence, the low-rank component $\hat{\mathbf{L}}$ can be ideally separated from the sparse changes and it will mainly

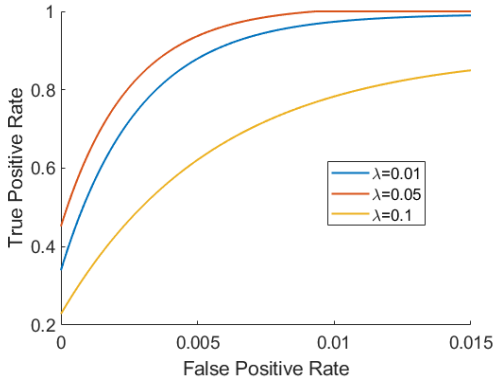


Figure 3.4: Receiving operating curves (fitted by an exponential function) for the winter night-time period showing the true positives versus false positives for detecting EV events.

capture the solar irradiance pattern. The recovery formulation is further tested on the winter night-time period from 18:00-24:00 to demonstrate the formulation's capability of identifying EV events in the recovered $\hat{\mathbf{P}}$. To this end, a receiving operating curve (ROC) is used to identify the accuracy of detected EV start/stop charging events. Fig. 3.4 plots the ROCs fitted with an exponential function for three λ values. For each value of λ, the ROC is generated by varying the detection threshold which is defined as a percentage of the EV's power rating. Note that the false positive rate is very small. This is because EVs have a much larger power rating than most other household appliances. As a result, the ROCs show a near perfect detection, and thus demonstrates the formulation's EV identification performance. In addition, the ROCs verify that the selected simulation value $\lambda = 0.05$ was a good choice. To further prove the effectiveness of the recovery formulation, one residential load is selected to demonstrate the recovery of an EV charging event occurring around 20:30-23:00. The selected residential load is shown in Fig. 3.5.

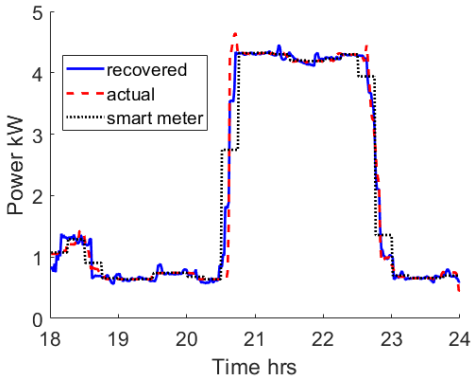


Figure 3.5: Recovered load data for one residential home with an EV charging event in the winter night-time period.

The smart meter measurements clearly fail to indicate the exact timing of the EV event. However, the estimated profile well matches that of the actual.

3.3.2 Summer Data Results

The effectiveness of Algorithm 1 is tested on the day-time period from 9:00-16:00 of the summer dataset. The night-time period is not tested here as the focus is to demonstrate the impact of periodic HVAC activities on the recovery solution. Compared to the previous winter cases, the summer data has a high level of HVAC activity. Fig. 3.1b shows the summer load profile of one residential home with HVAC activity. This periodic load manifests as a low-rank component. Therefore, it will not be fully contained in the sparse component $\mathbf{D}$ as a portion of the periodic magnitude will be captured by the low-rank component $\mathbf{K}$. To help decrease the periodic load impact on $\mathbf{K}$, the post-processing step in Algorithm 1 is applied to help improve recovery. The summer day-time ground-truth load profiles used in

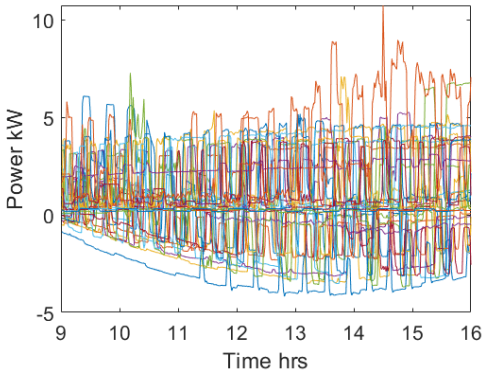
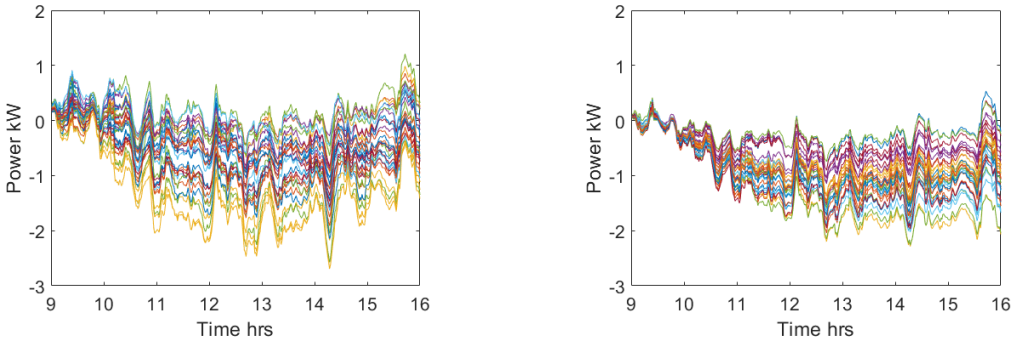


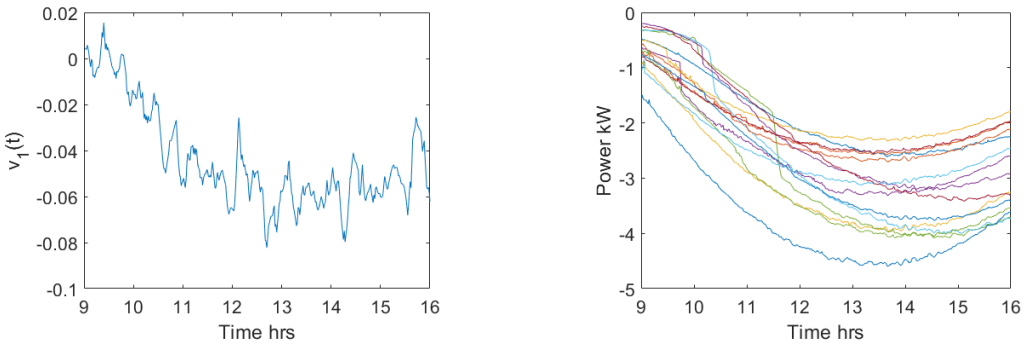
Figure 3.6: Ground-truth load profile for the summer day-time period.

this test case is shown in Fig. 3.6. The estimated $\hat{\mathbf{L}}$ using both (3.6) and Algorithm 1 is shown in Fig. 3.7. The overall estimation between the two estimations of $\hat{\mathbf{L}}$ is very similar. However, it can be observed that the post-processing step improved the low-rank recovery. Notice that in Fig. 3.7a there are large transients that appears to occur periodically. This is due to the low-rankness of the periodic HVAC and the magnitude bias of $\hat{\mathbf{D}}$ in (3.6). After running the post-processing step, it can be



(a) (b)

Figure 3.7: Recovered summer day-time $\hat{\mathbf{L}}$ solution from (a) formulation in (3.6) and (b) Algorithm 1.



(a) (b)

Figure 3.8: Solar irradiance pattern comparison between (a) the first singular vector of the estimated $\hat{\mathbf{L}}$ using Algorithm 1 and (b) the PV-only output data.

observed that the magnitude of the large transients is reduced in Fig. 3.7b. To better observe the solar irradiance pattern recovery, Fig. 3.8 compares the first singular vector of the estimated post-process $\hat{\mathbf{L}}$ and the ground-truth PV output profiles. The parabolic trend in the ground-truth PV output is well recovered, however, $\hat{\mathbf{L}}$ also exhibits some oscillatory patterns that occur around 30 minute intervals. This is due to the periodic HVAC activities seen in the ground-truth load profiles. Hence, the recovered solar irradiance pattern of the summer dataset is not as smooth as the actual PV profiles. In addition, compared to the winter day-time recovery results, the presence of HVAC activities degrades the recovered solution.

Chapter 4

Network Load Recovery Model

This chapter expands upon the simple model discussed in Chapter 3 and extends it to the network model. The distribution network presented previously in Chapter 2 is assumed to be a lossy multiphase system. The low-rank and sparse change characteristics of $\mathbf{P}$ exploited in Chapter 3 were shown to be effective in the numerical results of the simplified algorithm. The same two characteristics are exploited to recover $\mathbf{P}$ in the network algorithm presented later in this chapter. In addition, the underlying characteristics of the reactive power demand $\mathbf{Q}$ will be exploited in conjunction with characteristics of $\mathbf{P}$ in the algorithm to recover $\mathbf{P}$ and $\mathbf{Q}$.

4.1. Matrix Q Characteristics

Residential loads are known to be inductive, and thus have a lagging power factor (pf). This results from the majority of household appliances having a power factor less than 1 [40]. As such, there exists a relationship between the entries in real power demand $\mathbf{P}$ and the entries in reactive power demand $\mathbf{Q}$ such that $\mathsf{pf}_{n,t} = |p_{n,t}|/\sqrt{p_{n,t} + q_{n,t}}\ \forall (n,t)$. Therefore, the reactive power $\mathbf{Q}$ is given by

$$q_{n,t} = \frac{p_{n,t}}{\mathsf{pf}_{n,t}} \sqrt{1 - \mathsf{pf}_{n,t}^2}\ \forall (n,t). \tag{4.1}$$

Clearly, if a large household appliance were to turn on/off, there would a corresponding change in the reactive power for an increase/decrease in the amount of real power consumed. Therefore, the sparse changes in $\mathbf{Q}$ are *synchronized* with the sparse changes in $\mathbf{P}$. However, this is not the case with PVs. Recall from Chapter 3 that PVs contributed to the low-rank component in $\mathbf{P}$. The IEEE Standard 1547-2018 [11] recommends that distributed energy resources operate at unity power factor. Hence, it is assumed that the residential PV inverters operate at unity power factor, and thus $\mathbf{Q}$ does not have a low-rank component. As a result, $\mathbf{Q}$ only consists of sparse changes such that the consecutive differences in entries of $\mathbf{Q}$ are sparse, as defined by $D_{q_{n,t}} = Q_{n,t} - Q_{n,t-1} \forall (n,t)$ with $Q_{n,t} = \sum_{\tau=1}^{t} D_{q_{n,t}}$. In other words, $\mathbf{Q}$ becomes sparse under the linear transformation given by $\mathbf{Q} = \mathbf{D}_q \mathbf{U}$ where $\mathbf{U}$ is an upper triangular matrix of all ones.

4.2. Recovery Formulation

The decomposition of the real and reactive load matrices are given as $\mathbf{P} = (\mathbf{K} + \mathbf{D}_p)\mathbf{U}$ and $\mathbf{Q} = \mathbf{D}_q \mathbf{U}$ respectively. Clearly, the RPCA's low-rank plus sparse structure is maintained in both $\mathbf{P}$ and $\mathbf{Q}$. Relating the $\mathbf{K}$, $\mathbf{D}_p$, and $\mathbf{D}_q$ components back to the smart meter models in (2.1), the updated models are given as

$$\mathbf{\Gamma}_p = (\mathbf{K} + \mathbf{D}_p)\mathbf{U}\mathbf{A} + \mathbf{\Psi}_p \tag{4.2a}$$

$$\mathbf{\Gamma}_q = \mathbf{D}_q \mathbf{U}\mathbf{A} + \mathbf{\Psi}_q. \tag{4.2b}$$

Similarly, the real and reactive aggregated demand measurements, $\mathbf{Z}_p$ and $\mathbf{Z}_q$ respectively, are updated such that

$$\mathbf{X} \coloneqq -\begin{bmatrix} (\mathbf{K} + \mathbf{D}_p)\mathbf{U} \\ \mathbf{D}_q\mathbf{U} \end{bmatrix}. \tag{4.3}$$

The problem of recovering the load matrices $\mathbf{P}$ and $\mathbf{Q}$ becomes one of recovering a low-rank plus two sparse matrices. To promote the special structures of the load matrices, meaningful regularization terms are applied. The nuclear norm introduced in Chapter 3 is again used to promote the low-rank structure of $\mathbf{K}$. However, the L1-norm is not sufficient enough to promote both sparsity and synchronous non-zero values for corresponding entries in $\mathbf{D}_p$ and $\mathbf{D}_q$. Therefore, we introduce the group-Lasso regularization [19, 20] on $\mathbf{D}_p$ and $\mathbf{D}_q$ given by

$$\sum_{n,t} \|[D_{p_{n,t}}, D_{q_{n,t}}]\|_2. \tag{4.4}$$

Clearly, there are NT groups of size two in (4.4). In addition, note that the group-Lasso is essentially an L1-norm on the groups of variables as (4.4) reduces to the L1-norm if the groups are of size one. Hence, it promotes sparsity at the group level by promoting groups of variables to be either all non-zero or zero.

Utilizing the nuclear norm and group-Lasso regularizations, the matrices $\mathbf{P}$

and $\mathbf{Q}$ recovery problem can be defined as

$$\min_{\mathbf{K},\mathbf{D}_p,\mathbf{D}_q} \quad \|\mathbf{K}\|_* + \alpha \sum_{n,t} \|[D_{p_{n,t}}, D_{q_{n,t}}]\|_2 \tag{4.5a}$$

$$\text{s. to} \quad -\boldsymbol{\xi}_p \leq \boldsymbol{\Gamma}_p - (\mathbf{K} + \mathbf{D}_p)\mathbf{U}\mathbf{A} \leq \boldsymbol{\xi}_p \tag{4.5b}$$

$$-\boldsymbol{\xi}_q \leq \boldsymbol{\Gamma}_q - \mathbf{D}_q\mathbf{U}\mathbf{A} \leq \boldsymbol{\xi}_q \tag{4.5c}$$

$$-\boldsymbol{\phi}_p \leq \mathbf{Z}_p - \mathbf{H}_p\mathbf{X} \leq \boldsymbol{\phi}_p \tag{4.5d}$$

$$-\boldsymbol{\phi}_q \leq \mathbf{Z}_q - \mathbf{H}_q\mathbf{X} \leq \boldsymbol{\phi}_q \tag{4.5e}$$

$$\mathsf{vec}(\mathbf{D}_q\mathbf{U}) \geq \mathbf{0} \tag{4.5f}$$

where $\alpha > 0$ is a fixed weight coefficient to balance the matrices $\mathbf{K}$, $\mathbf{D}_p$, and $\mathbf{D}_q$ and the parameters $\boldsymbol{\xi}_p, \boldsymbol{\xi}_q, \boldsymbol{\phi}_p, \boldsymbol{\phi}_q > 0$ are pre-determined entry-wise error bounds. Note that $\mathbf{H}_p$ and $\mathbf{H}_q$ were defined in (2.9) and $\mathbf{X}$ was given by (4.3). In addition, note that the constraint (4.5f) results from the assumption that residential loads are inductive, and thus will only consume reactive power. Similar to Chapter 3, the error constraints in (4.5b)-(4.5e) correspond to uniformly distributed noise, and thus the error is bound using the infinity norm. Additionally, different error criteria can be used to replace (4.5b)-(4.5e) such as the Frobenius norm for Gaussian distributed noise. The recovery problem in (4.5) still has a convex objective and linear constraints and, therefore, it can be solved using generic convex optimization solvers such as CVX [34] and YALMIP [35].

4.3. Numerical Tests

Numerical studies using the recovery formulation in (4.5) to recover $\mathbf{P}$ and $\mathbf{Q}$ was conducted on a winter and summer dataset, namely the datasets used in

Chapter 3. Recall that the real power minute-level datasets were obtained from Pecan Street's Dataport [1] for 30 residential homes in the same neighborhood in Austin, TX. Additionally, only half of the houses have PVs and 6 of the houses have EV charging events. However, the reactive power data was not available from Pecan Street's Dataport. Therefore, for simplicity, the non-PV loads are assumed to have a constant power factor of 0.9. Hence, the reactive power demand $\mathbf{Q}$ was generated using (4.1) where $\mathsf{pf}_{n,t} = 0.9\forall(n,t)$ and $p_{n,t}\forall(n,t)$ is the real power demand without PV generation. To simulate a network model, a small residential feeder system was built in GridLab-D [41] by modifying GridLab-D's R2-12.47-3 taxonamy feeder. The 30 residential homes were distributed throughout the feeder such that there are 9, 9, and 12 houses connected to phases a, b, and c respectively in the multiphase system. The Pecan Street real power data and the synthetically generated reactive power data were used as input to the loading of the GridLab-D system and a simulation was run to record the aggregated demand at the feeder head. Note that the winter and summer data sets were scaled by times 15 and 6 respectively to match the orginal loading of the modified GridLab-D system. The recorded aggregated measurements are used to synthetically generate $\mathbf{Z}_p$ and $\mathbf{Z}_q$ using (2.9a) and (2.9b) respectively by adding random noise. Recall from Chapter 3, the D-PMU total vector error of $\pm 0.01\%$ [39] will result in a $\pm 0.02\%$ error in the real and reactive power measurements. Hence, the entries of the real and reactive aggregated measurement noise matrices $\mathbf{\Theta}_p$ and $\mathbf{\Theta}_q$ respectively are independently drawn from uniform distributions using $\pm 0.02\%$ accuracy. The real and reactive smart meter measurements were synthetically generated using (2.1a) and (2.1b)

respectively by adding random noise. Recall from Chapter 3, smart meters are assumed to be rated with $\pm 0.2\%$ accuracy from the ANSI C12.20 Standard [38]. Thus, the entries of the real and reactive smart meter measurement noise matrix $\mathbf{\Psi}_p$ and $\mathbf{\Psi}_q$ respectively are independently drawn from a uniform distribution based on this accuracy.

In addition to the error bounds, the group-Lasso weight coefficient α and the error bounds $\boldsymbol{\xi}_p$, $\boldsymbol{\xi}_q$, $\boldsymbol{\phi}_q$, and $\boldsymbol{\phi}_q$ need to be determined. In the following simulations, the weighting parameter α was set to a fixed value of 0.05 for the winter test case and a fixed value of 0.0913 for the summer test case. These values were chosen based upon the influenence of $\mathcal{O}(1/\sqrt{T})$ as discussed in Sec. 3.2.1 and trial and error to match the sparsity level of the estimated and original data. The error bounds $\boldsymbol{\xi}_p$, $\boldsymbol{\xi}_q$, $\boldsymbol{\phi}_q$, and $\boldsymbol{\phi}_q$ were determined based upon the metering accuracy used to generate the synthetic data. Hence, $\boldsymbol{\xi}_p$, $\boldsymbol{\xi}_q$, $\boldsymbol{\phi}_q$, and $\boldsymbol{\phi}_q$ are set to $0.2\%|\mathbf{\Gamma}_p|$, $0.2\%|\mathbf{\Gamma}_q|$, $0.02\%|\mathbf{Z}_p|$, and $0.02\%|\mathbf{Z}_q|$ respectively.

Last, the matrices $\mathbf{H}_p$ and $\mathbf{H}_q$ in (2.9a) and (2.9b) are pre-calculated. Recall that $\mathbf{H}_p = \mathsf{Re}\{\mathbf{H}\}$, $\mathbf{H}_q = \mathsf{Im}\{\mathbf{H}\}$, and $\mathbf{H} = \mathsf{diag}(\mathbf{v}_0)\mathbf{Y}_{0L}^{*}\mathbf{M}^{*}$ where $\mathbf{M}$ is defined by (2.6) for a given input voltage profile. To obtain a better estimate of $\mathbf{M}$ for a time period of interest, we calculate the system voltage $\tilde{\mathbf{v}}$ for the average power injection at each PQ bus by using the fixed point approximation method in (2.5) where $\mathbf{x}$ consists of the averaged real and reactive smart meter data. The estimated $\tilde{\mathbf{v}}$ is plugged into (2.6) to estimate $\mathbf{M}$ which is used to determine $\mathbf{H}$.

Note that due to the increase in the number of unknown variables in (4.5), MATLAB [42] only has enough memory to simulate two hour time periods. As a

result, only two hours will be simulated for the winter day-time and summer test cases. However, the for the winter night-time test case, (4.5) was run three times to recover the load demand profiles for the six hours of interest.

4.3.1 Winter Data Results

The multiphase recovery formulation (4.5) is tested on two time periods in the winter data set, namely one day-time period and one night-time period. Recall

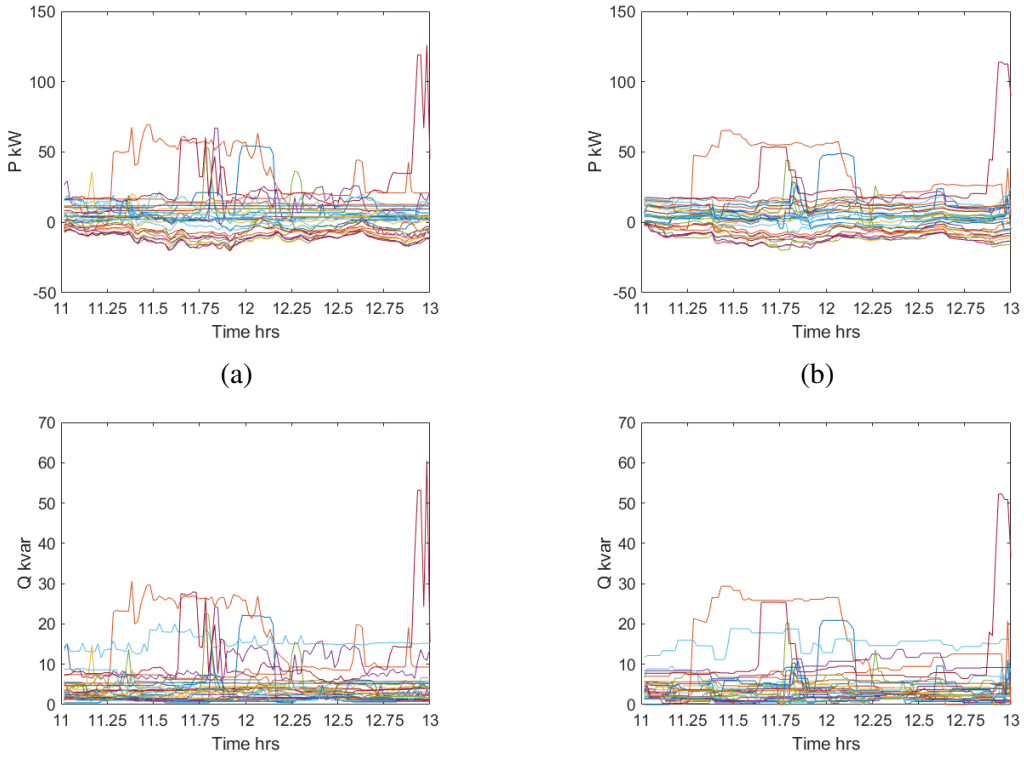


Figure 4.1: Comparison between the real and reactive ground-truth residential load profiles, (a) and (c) respectively, and the recovered $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$, (b) and (d) respectively, from (4.5) for the winter day-time period.

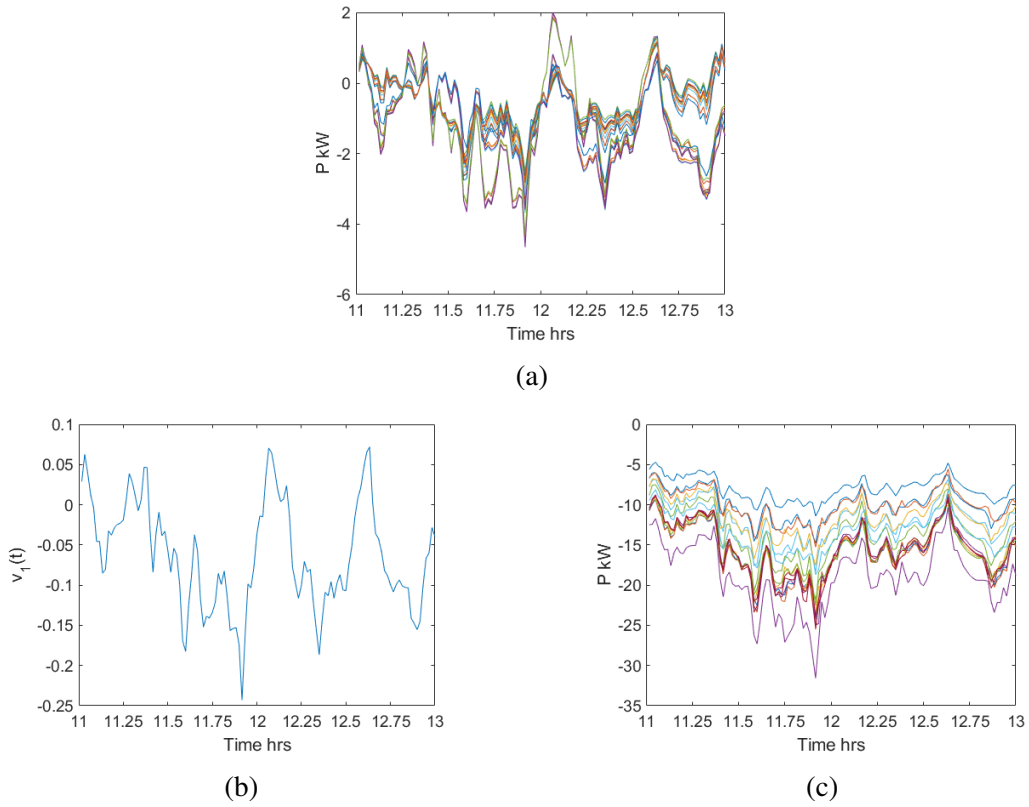


Figure 4.2: Comparison of the winter day-time solar irradiance pattern recovery of (a) the estimated $\hat{\mathbf{L}}$ and (b) its first right singular vector with (c) the ground-truth PV output profile.

that the winter dataset was scaled by times 15 to better match the original GridLab-D system loading. The day-time period from 11:00-13:00 was tested first to recover the solar irradiance pattern. Starting with the overall recovery of $\mathbf{P}$ and $\mathbf{Q}$, Fig. 4.1b and 4.1d plots the estimated $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ respectively which well matches the trends in the real and reactive ground-truth profiles shown in Fig. 4.1a and 4.1c respectively. Although neither $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$ can recover the fast transients, they are capable of capturing major dynamic changes in the load profiles such as large appliance

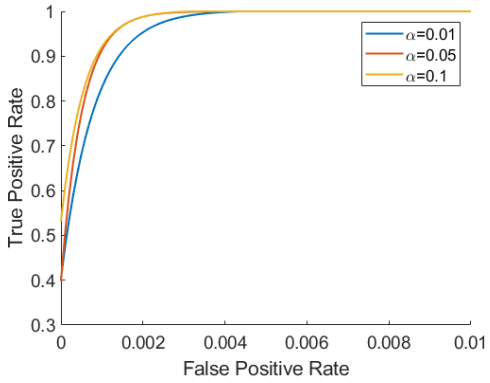


Figure 4.3: Receiving operating curves (fitted by an exponential function) for the winter night-time period showing the true positives versus false positives for detecting EV events.

events. To better observe the solar irradiance pattern recovery, Fig. 4.2 shows the comparison between the estimated $\hat{\mathbf{L}} = \hat{\mathbf{K}}\mathbf{U}$ and the ground-truth PV output profiles. Note that the estimated $\hat{\mathbf{L}}$ shown in Fig. 4.2a does not recover the actual PV output shown in Fig. 4.2c. However, it is clear that the estimated $\hat{\mathbf{L}}$ is of very low rank. In addition, its first right singular vector shown in Fig. 4.2b captures not only most of its own temporal pattern, but also the solar irradiance pattern present in the ground-truth PV outputs. Similar to the observations made in Chapter 3, the winter dataset contains no HVAC activities compared to the summer dataset. Therefore, the low-rank component $\hat{\mathbf{L}}$ can be well separated from the sparse changes and it will mainly capture the solar irradiance pattern. The multiphase recovery formulation was further tested on the winter night-time period from 18:00-24:00 to demonstrate the formulation's capability of recovering EV events in the recovered $\hat{\mathbf{P}}$. To this end, a ROC plot is used to identify the accuracy of detected EV start/stop charging events. Fig. 4.3 plots the ROCs fitted with an exponential function for three α val-

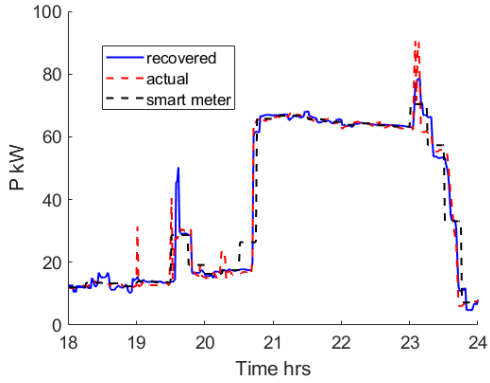
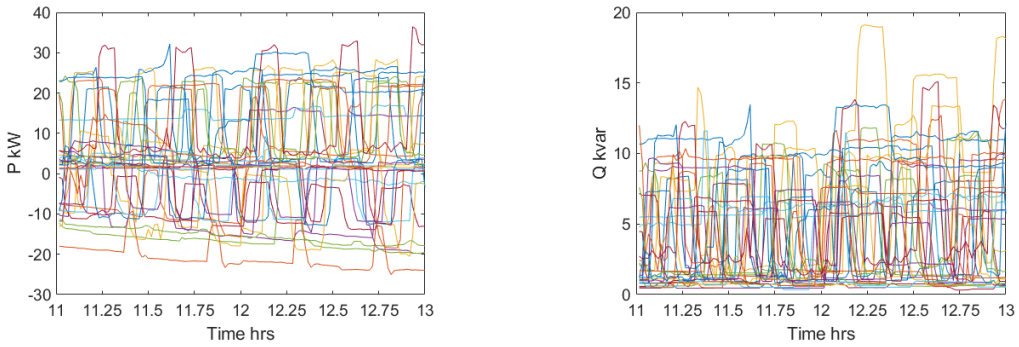


Figure 4.4: Recovered load profile for one residential home with an EV charging event in the winter night-time period.

ues. Recall from Chapter 3, the ROC is generated by varying the detection threshold for each α value. Similar to Chapter 3, the false positive rate is very small as EV's have larger power ratings than most household appliances. Hence, the ROC results show a near perfect detection and, thus, demonstrates the multiphase formulation's EV identification performance. Furthermore, the ROCs verify that the value 0.05 was a good choice for the group-Lasso weight parameter α. To further show the effectiveness of the recovery formulation, Fig. 4.4 shows one recovered residential load profile to demonstrate the recovery of an EV charging event occurring around 20:40-23:45. Clearly, the recovered profile well matches the actual while the smart meter measurements fail to capture the exact timing of the EV event.

4.3.2 Summer Data Results

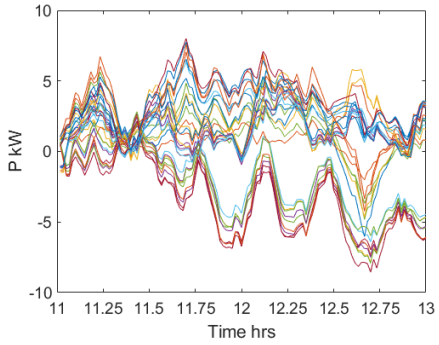
The effectiveness of the multiphase recovery formulation (4.5) is tested on the day-time period from 11:00-13:00 of the summer dataset. The night-time period



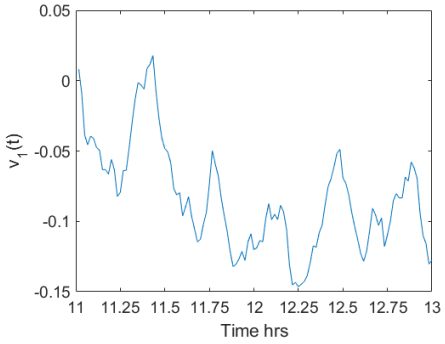
(a) (b)

Figure 4.5: Ground-truth load profiles for the real and reactive power demand for the summer day-time period.

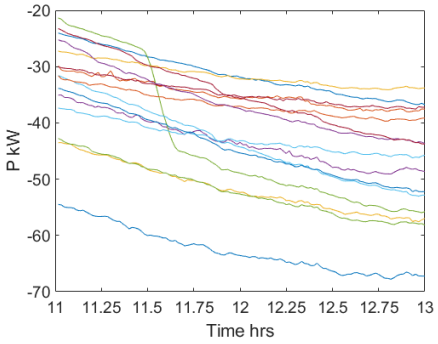
is not tested here as the focus is to demonstrate the impact of periodic HVAC activities on the recovery solution. Recall that the summer dataset was scaled by times 6 to better match the original GridLab-D system loading. The summer dataset has a high level of HVAC activity as shown in the ground-truth load profiles in Fig. 4.5. Therefore, the sparse component $\mathbf{D}_p$ will not fully capture the HVAC events as a portion of the periodic magnitude will be contained in the low-rank component $\mathbf{K}$. This effect is visible in the estimated $\hat{\mathbf{L}}$ shown in Fig. 4.6a where there are large transients that occur periodically. The periodicity in $\hat{\mathbf{L}}$ is more clearly observed in the right singular vector shown in Fig. 4.6b where the oscillatory patterns occur at approximately 20 minute intervals. Although $\hat{\mathbf{L}}$ captures a portion of the HVAC activities, it is capable of capturing the solar irradiance pattern trend. Comparing the first right singular vector in Fig. 4.6b to the ground-truth PV output in Fig. 4.6c, the slight parabolic trend is recovered. However, compared to the winter solar irradiance recovery, the summer recovery results as not as smooth as the actual PV



(a)



(b)



(c)

Figure 4.6: Comparison of the summer day-time solar irradiance pattern recovery of (a) the estimated $\hat{\mathbf{L}}$ and (b) its first right singular vector with (c) the ground-truth PV output profile.

outputs due to the high level of HVAC activities. Clearly, the presence of HVAC degrades the recovered solution.

Chapter 5

Improving the Recovery of Periodic Loads

This chapter presents an additional sinusoidal component to address the recovery result degradation caused by the presence of periodic HVAC activities. It is well known in the signals field that periodic signals can be decomposed into the summation of sines and cosines. In addition, the signal can be assumed to have a maximum rank of two if only the first harmonic is considered. To this end, the simplified and network models are augmented with an additional low-rank sinusoidal component. Numerical studies using the summer dataset were conducted on the new simplified and network recovery formulations.

5.1. Simplified Load Recovery Model

As observed in the summer simulation results in Chapter 3, periodic HVAC activities exhibit an inherent low-rankness which degrades the recovery performance. To alleviate the impact of HVAC bias in $\mathbf{L}$, it is possible to introduce an additional sinusoidal component to capture the periodic patterns. It is well known in the signals field that a periodic signal $g(t)$ can be decomposed as the summation

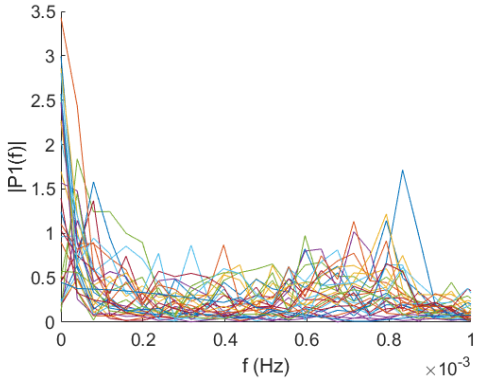


Figure 5.1: Spectral analysis results of each ground-truth residential load profile.

of sines and cosines such that

$$g(t) = \sum_{k=1}^{\infty} a_k \sin(k\omega_0 t) + b_k \cos(k\omega_0 t) \tag{5.1}$$

where k represents the k-th harmonic component and ω_0 is the fundamental frequency. For simplicity, $g(t)$ is assumed to consist of only the first harmonic component, such that $k = 1$. Hence, the collection of such periodic signals, all with frequency ω_0, can be defined in matrix form as

$$\mathbf{G} = \begin{bmatrix} a_1 & b_1 \\ \vdots & \vdots \\ a_N & b_N \end{bmatrix} \begin{bmatrix} \sin(\omega_0 t_1) & \sin(\omega_0 t_2) & \cdots & \sin(\omega_0 T) \\ \cos(\omega_0 t_1) & \cos(\omega_0 t_2) & \cdots & \cos(\omega_0 T) \end{bmatrix} \tag{5.2}$$

$$= \mathbf{CF} \tag{5.3}$$

where $\mathbf{C} \in \mathbb{R}^{N \times 2}$ is the matrix of coefficients and $\mathbf{F} \in \mathbb{R}^{2 \times T}$ is the matrix of sines and cosines. Clearly, the periodic component $\mathbf{G}$ has at most rank 2, which corroborates the low-rankness of HVAC activities and its impact on the recovery of

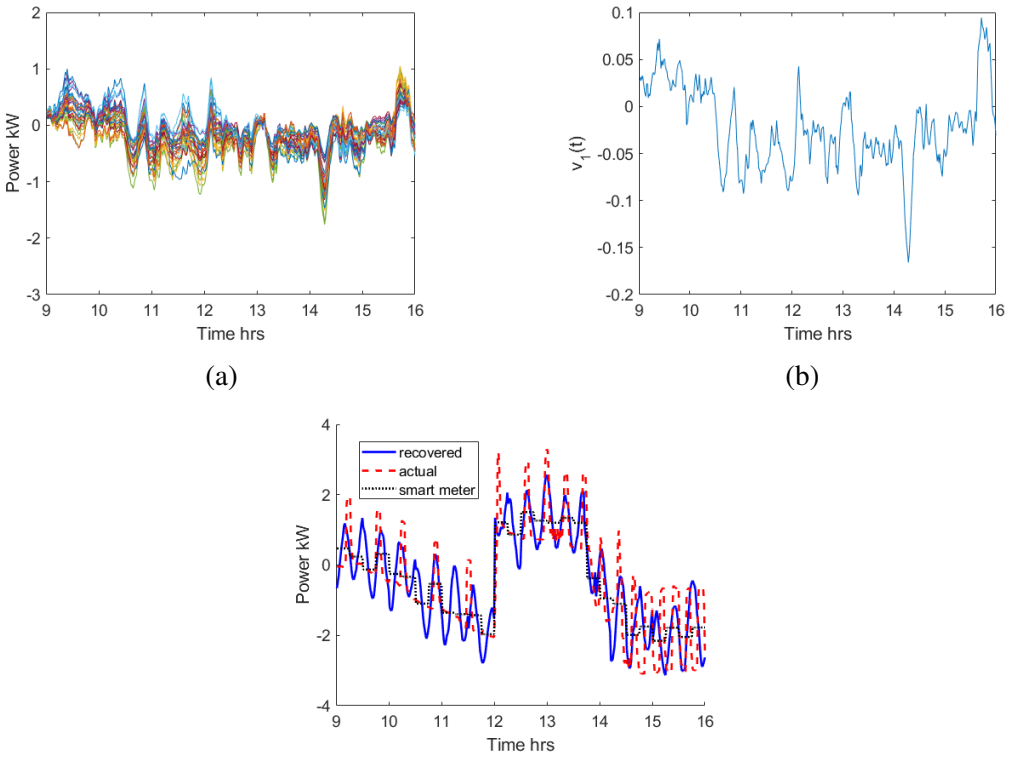


Figure 5.2: Recovery results of (5.5) on the summer dataset showing (a) the estimated $\hat{\mathbf{L}}$,(b) the first right singular vector, and (c) the recovered load profile of a selected house.

$\mathbf{L}$ and $\mathbf{D}$. Augmenting the new sinusoidal component into the system model, $\mathbf{P}$ is now decomposed as $\mathbf{P} = \mathbf{L} + \mathbf{S} + \mathbf{G} = \mathbf{L} + \mathbf{DU} + \mathbf{CF}$ and (3.2) becomes

$$\mathbf{R} = \mathbf{PW} = \mathbf{K} + \mathbf{D} + \mathbf{CJ} \tag{5.4}$$

where $\mathbf{J} = \mathbf{FW}$. As a result, the convex problem (3.6) now becomes

$$\min_{\mathbf{K},\mathbf{D},\mathbf{C}} \quad \|\mathbf{K}\|_* + \lambda\|\mathbf{D}\|_1 \tag{5.5a}$$

$$\text{s. to} \quad -\boldsymbol{\xi} \leq \boldsymbol{\Gamma}_p - (\mathbf{K} + \mathbf{D} + \mathbf{CJ})\mathbf{UA} \leq \boldsymbol{\xi} \tag{5.5b}$$

$$-\boldsymbol{\phi} \leq \mathbf{z}_p^{\mathsf{T}}\mathbf{W} - \mathbf{1}^{\mathsf{T}}(\mathbf{K} + \mathbf{D} + \mathbf{CJ}) \leq \boldsymbol{\phi} \tag{5.5c}$$

with the updated smart meter and aggregated demand models $\boldsymbol{\Gamma}_p = (\mathbf{K} + \mathbf{D} + \mathbf{CJ})\mathbf{UA}$ and $\mathbf{z}_p^{\mathsf{T}}\mathbf{W} = \mathbf{1}^{\mathsf{T}}(\mathbf{K} + \mathbf{D} + \mathbf{CJ})$ respectively.

To test the effectiveness of introducing a sinusoidal component, the new convex problem (5.5) is tested on the same summer day-time dataset used in Sec. 3.3.2. As shown in Fig. 5.1, the fundamental frequency, $f_0 \approx 0.8 \times 10^{-3}$ Hz, is obtained by performing spectral analysis on the ground-truth load profiles in Fig. 3.6. It is then used to define the matrix $\mathbf{F}$ with $\omega_0 = 2\pi f_0$. Fig. 5.2 shows the estimated $\hat{\mathbf{L}}$, the right singular vector, and the recovered load profile of one selected residential home. The recovered solar irradiance pattern shows no visible improvement from the previously recovered pattern in Fig. 3.8. However, upon closer inspection of the individual load profiles, the sinusoidal component is indeed attempting to capture the periodic HVAC activities.

5.2. Network Load Recovery Model

Similarly, the new multiphase formulation is obtained by augmenting the sinusoidal component to both the real and reactive power demands $\mathbf{P}$ and $\mathbf{Q}$ re-

spectively. As a result, the decomposition of $\mathbf{P}$ and $\mathbf{Q}$ is given as

$$\mathbf{P} = (\mathbf{K} + \mathbf{D}_p)\mathbf{U} + \mathbf{C}_p\mathbf{F} \tag{5.6a}$$

$$\mathbf{Q} = \mathbf{D}_q\mathbf{U} + \mathbf{C}_q\mathbf{F}. \tag{5.6b}$$

Hence, the multiphase recovery formulation now becomes

$$\min_{\mathbf{K},\mathbf{D}_p,\mathbf{D}_q,\mathbf{C}_p,\mathbf{C}_q} \quad \|\mathbf{K}\|_* + \alpha \sum_{n,t} \|[D_{p_{n,t}}, D_{q_{n,t}}]\|_2 + \beta(\|\mathbf{C}_p\|_F^2 + \|\mathbf{C}_q\|_F^2) \tag{5.7a}$$

$$\text{s. to} \quad -\boldsymbol{\xi}_p \leq \boldsymbol{\Gamma}_p - ((\mathbf{K} + \mathbf{D}_p)\mathbf{U} + \mathbf{C}_p\mathbf{F})\mathbf{A} \leq \boldsymbol{\xi}_p \tag{5.7b}$$

$$-\boldsymbol{\xi}_q \leq \boldsymbol{\Gamma}_q - (\mathbf{D}_q\mathbf{U} + \mathbf{C}_q\mathbf{F})\mathbf{A} \leq \boldsymbol{\xi}_q \tag{5.7c}$$

$$-\boldsymbol{\phi}_p \leq \mathbf{Z}_p - \mathbf{H}_p\mathbf{X} \leq \boldsymbol{\phi}_p \tag{5.7d}$$

$$-\boldsymbol{\phi}_q \leq \mathbf{Z}_q - \mathbf{H}_q\mathbf{X} \leq \boldsymbol{\phi}_q \tag{5.7e}$$

$$\mathsf{vec}(\mathbf{D}_q\mathbf{U} + \mathbf{C}_q\mathbf{F}) \geq \mathbf{0} \tag{5.7f}$$

where

$$\mathbf{X} \coloneqq -\begin{bmatrix} (\mathbf{K} + \mathbf{D}_p)\mathbf{U} + \mathbf{C}_p\mathbf{F} \\ \mathbf{D}_q\mathbf{U} + \mathbf{C}_q\mathbf{F} \end{bmatrix}, \tag{5.8}$$

β is the weight parameter for the sinusoidal components, and $\boldsymbol{\xi}_p, \boldsymbol{\xi}_q, \boldsymbol{\phi}_p, \boldsymbol{\phi}_q > 0$ are the entry-wise error bounds. The weighting parameter β was introduced to adjust the sparsity level of the recovered results.

The effectiveness of the new multiphase formulation (5.7) was tested on the same summer dataset. The sinusoidal weight parameter β was determined by tuning the parameter through trial and error. A fixed value of 0.005 was chosen for β in the following simulation results. Fig. 5.3 shows the estimated $\hat{\mathbf{L}}$, the first right singular

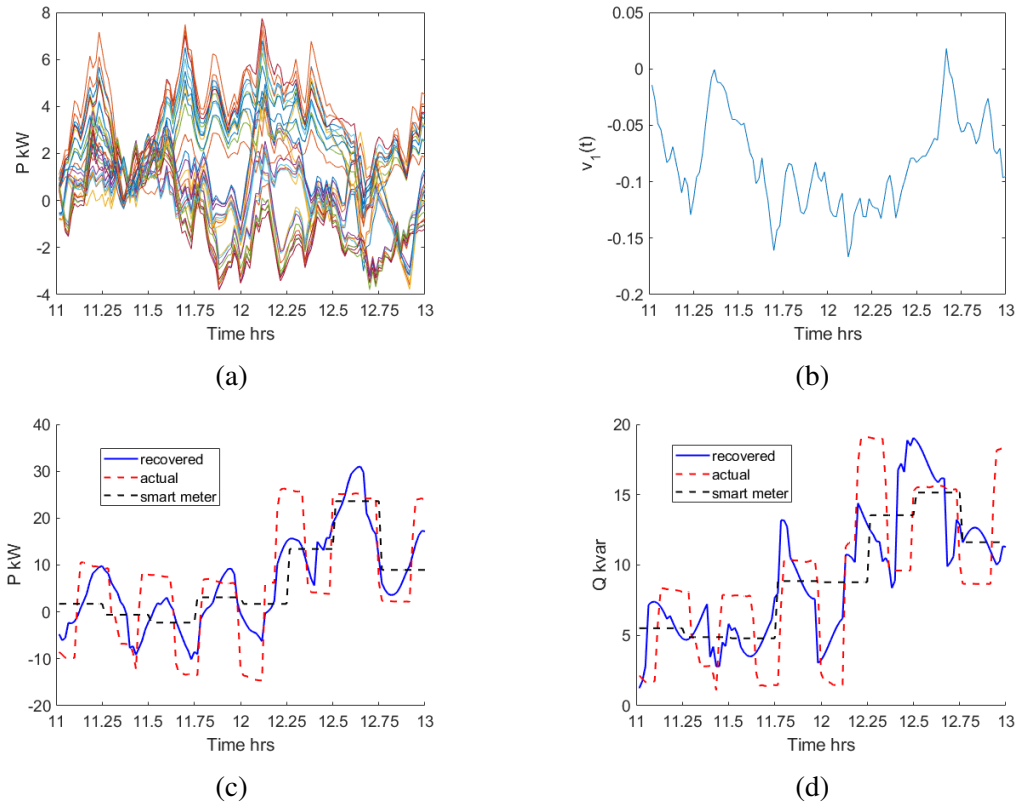


Figure 5.3: Recovery results of (5.7) on the summer dataset showing (a) the estimated $\hat{\mathbf{L}}$,(b) the first right singular vector, and the real and reactive load profiles of a selected house in (c) and (d) respectively.

vector, and the real and reactive load profiles of a selected house. Clearly, the estimated $\hat{\mathbf{L}}$ still contains an oscillatory pattern. Hence, the recovered solar irradiance pattern shows no visible improvement from the previously recovered pattern shown in Fig. 4.6b. However, the sinusoidal component is indeed trying to capture the real and reactive periodic HVAC activities shown in Fig. 5.3c and 5.3d respectively.

Chapter 6

Conclusions and Future Work

This thesis presented a spatio-temporal load recovery formulation for a simplified lossless model and a lossy multiphase network. The simplified model was utilized to show the effectiveness of exploiting the underlying low-rank and sparse change characteristics of the real power demand $\mathbf{P}$ in load recovery. The nuclear norm and L1-norm regularization were adopted to promote the low-rank and sparse components respectively. The simplified formulation was then extended to the multiphase network and the characteristics of the reactive power demand $\mathbf{Q}$ were introduced. It it assumed that PV inverters are of unity power factor and, therefore, $\mathbf{Q}$ does not have a low-rank component. Hence, $\mathbf{Q}$ contains only sparse changes that are synchronized with $\mathbf{P}$. As a result, the nuclear norm is again used to promote the low-rank component in $\mathbf{P}$. However, instead of the L1-norm, the group-LASSO regularization is used to promote synchronized changes and group level sparsity in the sparse components of $\mathbf{P}$ and $\mathbf{Q}$. The simplified and network recovery formulations were tested on a winter and a summer dataset. The solar irradiance patterns were well recovered for the winter day-time periods and the EV charging events were well identified in the winter night-time periods. Compared to the winter results, the summer recovery results were degraded due to the presence of high periodic HVAV activities. Although the summer solar irradiance pattern recovery

is capable of capturing the trend of the ground-truth PV output, it also captures a portion of the oscillatory pattern of HVAC activities.

To address the periodic challenge, an additional sinusoidal component was augmented to the recovery formulations as any periodic signal can be decomposed as the summation of sines and cosines. The fundamental frequency ω_0 was approximated by performing spectral analysis on the summer dataset to setup $\mathbf{F}$ the matrix of sines and cosines. The new recovery formulations with the sinusoidal component were tested using the summer dataset. Clearly, there is no visible improvement on the solar irradiance pattern recovery. However, upon closer inspection of the individual loads, the sinusoidal component is indeed attempting to capture the periodic HVAC activities.

Two future work directions are proposed in the following sections. One direction discusses the online implementations of the recovery algorithms to accelerate the computational speed and to process data streams in real-time. The other future direction discusses the use of a rectangular waveform model to better represent the presence of periodic HVAC loads.

6.1. Online Algorithm Implementations

Clearly, the computational complexity of the load recovery formulations presented in Chapters 3 and 4 will grow with the matrix dimension. This issue can be overcome by utilizing accelerated algorithms such as alternating minimization [43, 44] or adaptive updates using subspace tracking approaches [45–47]. The alternating minimization methods allows the low-rank component to be fac-

tored into two lower-dimensional matrices such that $\mathbf{K} = \mathbf{\Sigma\Omega}$ where $\mathbf{\Sigma} \in \mathbb{R}^{N\times r}$, $\mathbf{\Omega} \in \mathbb{R}^{r\times T}$, and r is the desired rank of $\mathbf{K}$. Clearly, the number of unknown variables will be reduced, however, the bi-linear parameterization would make the recovery problem non-convex. By using alternating minimization, $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ can be efficiently updated by iteratively solving the corresponding convex sub-problems. Although the alternative minimization method can speed up computation speed, the global optimality guarantees are lacking [44]. In addition, for a time-varying system, solving the formulation every time new measurements arrive results in redundant computation [48]. One can avoid redundant computation by using adaptive subspace tracking methods [45–47]. The subspace spanned by the columns of $\mathbf{\Sigma}$ is used to calculate the projection $\mathbf{\Omega}$ for a new measurement. In addition, the new measurement is used to update the subspace of $\mathbf{\Sigma}$. Hence, the adaptive subspace tracking method can be solved in real-time when new measurements are available due to fast computational speeds.

6.2. Rectangular Waveform Modeling for Periodic Loads

The results in Chapter 5 showed that the sinusoidal component recovery was not very good. For some residential houses, the periodic HVAC activities were well estimated. However, the recovered results does not fully capture the rectangular shape of the HVAC events. In addition, the sinusoidal component had little to no impact on the solar irradiance pattern recovery. To address this issue, one could cast the HVAC load as a rectangular wave with a fixed period. As a result, the goal will be to estimate the wave amplitude and duty cycle for every residential house.

Intuitively, the rectangular waveform better fits the HVAC load profile as HVAC units rarely remain on for half of its period. Furthermore, the duty cycle recovery will potentially allow for improved load profile recovery for each residential home. As a result, this formulation could lead to less oscillatory patterns in the low-rank component, and thus a better solar irradiance pattern recovery.

Bibliography

[1] PecanStreet Dataport.

[2] Anggoro Primadianto and Chan-Nan Lu. A Review on Distribution System State Estimation. *IEEE Transactions on Power Systems*, 32(5):3875–3883, Sept. 2017.

[3] C.N. Lu, J. H. Teng, and W.-H. E. Liu. Distribution system state estimation. *IEEE Transactions on Power Systems*, 10(1):229–240, Feb. 1995.

[4] Mesut E. Baran and Arthur W. Kelley. State estimation for real-time monitoring of distribution systems. *IEEE Transactions on Power Systems*, 9(3):1601–1609, Aug. 1994.

[5] Alexandra von Meier, David Culler, Alex McEachern, and Reza Arghandeh. Micro-Synchrophasors for Distribution Systems. *IEEE PES Innovative Smart Grid Tech. Conf.*, Feb. 2014.

[6] Advanced Metering Infrastructure and Customer Systems, Sept. 2016.

[7] Chaitra Niddodi, Shanny Lin, Sibin Mohan, and Hao Zhu. Secure integration of electric vehicles with the power grid. *arXiv preprint arXiv:1905.01035*, 2019.

[8] Nicholas Jacobs, Shamina Hossain-McKenzei, Deepu Jose, Danish Saleem, Christine Lai, Patricia Cordeira, Adarsh Hasandka, Maurice Martin, and Christopher Howerter. Analysis of System and Interoperability Impact from Securing Communications for Distributed Energy Resources. *2019 IEEE Power and Energy Conference at Illinois (PECI)*, 2019.

[9] Xiangtian Zheng, Xinbo Geng, Le Xie, Dongliang Duan, Liuqing Yang, and Shuguang Cui. A SVM-based setting of protection relays in distribution systems. *2018 IEEE Texas Power and Energy Conference (TPEC)*, 2018.

[10] S. Bhela, V. Kekatos, and S. Veeramachaneni. Enhancing observability in distribution grids using smart meter data. *IEEE Transactions on Smart Grid*, 9(6):5953–5961, 2018.

[11] Ieee standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces. *IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003)*, pages 1–138, 2018.

[12] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[13] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust PCA via Outlier Pursuit. *IEEE Transactions on Information Theory*, 58(5):3047–3064, 2012.

[14] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matri-

ces via convex optimization. In *Advances in Neural Information Processing Systems 22*, pages 2080–2088, 2009.

[15] Pengzhi Gao, Meng Wang, Scott G. Ghiocel, Joe H. Chow, Bruce Fardanesh, and George Stefopoulos. Missing Data Recovery by Exploiting Low-dimensionality in Power Systems Synchrophasor Measurements. *IEEE Trans. Power Systems*, 31(2):1006–1013, Mar. 2016.

[16] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang. Matrix completion for low-observability voltage estimation. *IEEE Transactions on Smart Grid*, 11(3):2520–2530, 2020.

[17] Gonzalo Mateos and Georgios B. Giannakis. Load Curve Data Cleansing and Imputation Via Sparsity and Low Rank. *IEEE Trans. on Smart Grid*, 4(4), Dec. 2013.

[18] Tong Huang, Nikolaos M. Freris, P. R. Kumar, and Le Xie. Localization of forced oscillations in the power grid under resonance conditions. *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018.

[19] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages 49–67, 2006.

[20] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. A note on the group lasso and a sparse group lasso. *arXiv preprint arXiv:1001.0736*, 2010.

[21] Andrey Bernstein, Cong Wang, Emiliano Dall'Anese, Jean-Yves Le Boudec, and Changhong Zhao. Load-Flow in Multiphase Distribution Networks: Existence, Uniqueness, Non-Singularity and Linear Models. *IEEE Trans. Power Sys.*, 33(6):5832–5843, 2018.

[22] William H. Kersting. *Distribution System Modeling and Analysis*. CRC Press, Boca Ranton, FL, 2 edition, 7007.

[23] Saverio Bolognani and Florian Dörfler. Fast power system analysis via implicit linearization of the power flow manifold. *53rd Annual Allerton Conference on Communiation, Control, and Computing*, 2015.

[24] Sairaj V. Dhople, Swaroop S. Guggilam, and Yu Christine Chen. Linear approximations to AC power flow in rectangular coordinates. *53rd Annual Allerton Conference on Communiation, Control, and Computing*, 2015.

[25] M.E. Baran and F.F. Wu. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power Delivery*, 4(2), Apr. 1989.

[26] Pu Wang, Bidong Liu, and Tao Hong. Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting*, 32(3):585–597, 2016.

[27] Jean-Baptiste Fiot and Francesco Dinuzzo. Electricity Demand Forecasting by Multi-Task Learning. *IEEE Trans. on Smart Grid*, 9(2):544–551, Mar. 2018.

[28] Seung-Jun Kim and Geogios B. Giannakis. Load Forecasting via Low Rank Plus Sparse Matrix Factorization. *2013 Asilomar Conference on Signals, Systems and Computers*, Nov. 2013.

[29] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[30] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, 1st edition, 2004.

[31] EJ Candes and T Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 12(52):5406–5425, 2006.

[32] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[33] David L. Donoho. For Most Large Underdetermined Systems of Linear Equations of Minimal L1-norm Solution Is Also The Sparsest Solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, 2006.

[34] CVX: Matlab Software for Disciplined Convex Programming, 2018.

[35] J. Löfberg. Yalmip : A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[36] Namrata Vaswani, Thierry Bouwmans, Sajid Javed, and Praneeth Narayanamurthy. Robust subspace learning: Robust pca, robust subspace tracking, and robust subspace recovery. *IEEE Signal Processing Magazine*, 35(4):32–55, 2018.

[37] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. Rank-Sparsity Incoherence for Matrix Decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[38] ANSI C12.20-2015-Electricity Meters-0.1, 0.2, and 0.5 Accuracy Classes, 2015.

[39] microPMU Data Sheet, 2018.

[40] Manisa Pipattanasomporn, Murat Kuzlu, Saifur Rahman, and Yonael Teklu. Load Profiles of Selected Major Household Appliances and Their Demand Response Opportunities. *IEEE TRANSACTIONS ON SMART GRID*, 5(2), 2014.

[41] GridLab-D.

[42] MATLAB.

[43] Moritz Hardt. Understanding alternating minimization for matrix completion. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 651–660. IEEE, 2014.

[44] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674, 2013.

[45] L. Balzano, R. Nowak, and B. Recht. Online identification and tracking of subspaces from highly incomplete information. In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 704–711, 2010.

[46] Jun He, Laura Balzano, and John C.S. Lui. Online robust subspace tracking from partial information. *arXiv preprint arXiv:1109.3827*, 2011.

[47] Morteza Mardani, Gonzalo Mateos, and Georgios B Giannakis. Dynamic anomalography: Tracking network anomalies via sparsity and low rank. *IEEE Journal of Selected Topics in Signal Processing*, 7(1):50–66, 2012.

[48] U. Niesen, D. Shah, and G. Wornell. Adaptive alternating minimization algorithms. In *2007 IEEE International Symposium on Information Theory*, pages 1641–1645, 2007.

Vita

Shanny Lin was born in El Paso, Texas. After completing her curriculum at El Paso High School, El Paso, Texas in 2014, she was accepted at Rensselaer Polytechnic Institute in Troy, New York as a varsity student-athlete. She received the Bachelor of Science degree in Electrical Engineering from Rensselaer Polytechnic Institute in May 2018. Following her undergraduate studies, she was accepted into the Graduate School at the University of Texas at Austin in August, 2018.

E-mail address: shannylin@live.com

This thesis was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.